

# **Global AI Literacy Study**

Dr. Anastassia Lauterbach (ORCID ID 0009-0006-6988-7406)

September, 2026

|                                                                                                     |           |
|-----------------------------------------------------------------------------------------------------|-----------|
| <b>ABSTRACT .....</b>                                                                               | <b>5</b>  |
| <b>INDEPENDENCE, CONTRIBUTORS AND USE OF AI.....</b>                                                | <b>7</b>  |
| <b>DISCLAIMER.....</b>                                                                              | <b>8</b>  |
| <b>CHAPTER 1. DEFINING AI LITERACY AND AI FLUENCY .....</b>                                         | <b>9</b>  |
| A MULTI-DIMENSIONAL MEANING OF “LITERACY” .....                                                     | 9         |
| AI LITERACY VS. AI FLUENCY .....                                                                    | 9         |
| THE STAKES: WHAT ILLITERACY MEANS FOR BUSINESSES AND EMPLOYMENT.....                                | 11        |
| “AI EXPERTISE” .....                                                                                | 12        |
| SOURCES TO CHAPTER 1 .....                                                                          | 15        |
| <b>CHAPTER 2: UNDERSTANDING AI, ITS RISKS, IMPACT AND COSTS.....</b>                                | <b>18</b> |
| DEFINITIONS OF AI AND DIFFERENCES TO HUMAN LEARNING .....                                           | 18        |
| AI RISKS .....                                                                                      | 20        |
| <i>Risk 1 - The structural risk: bias baked in, and models that are grown, not built</i><br>.....   | 20        |
| <i>Workarounds Instead of New Architectures: Tooling, MCP and Context Rot ...</i>                   | 22        |
| <i>Risk 2 - The (malicious) use risk: (bad) intent meets capability .....</i>                       | 24        |
| <i>Risk 3 - The calibration risk: humans in, on, or out of the loop.....</i>                        | 26        |
| <i>Specifications, Prompts and Inputs: When AI Does What We Said, Not What We Meant.....</i>        | 29        |
| <i>Risk 4 - The Overreliance Risk/ “Cognitive Debt” .....</i>                                       | 32        |
| <i>Risk 5 - The Catastrophic Risk.....</i>                                                          | 33        |
| ADDRESSING AI RISKS .....                                                                           | 33        |
| <i>Enforcing Radical Honesty about AI Capabilities, Bottlenecks and Use of Synthetic Data .....</i> | 34        |
| <i>Establishing Vocabulary to Discuss AI .....</i>                                                  | 35        |
| DECIDING ON AI IMPACT AND DEVELOPING AN AI STRATEGY.....                                            | 36        |
| AI IMPLEMENTATION COSTS .....                                                                       | 40        |
| <i>Model-related Pricing and the Harness .....</i>                                                  | 42        |
| <i>Sovereignty.....</i>                                                                             | 49        |
| <i>Geography/ compliance regime.....</i>                                                            | 52        |
| <i>Hardware .....</i>                                                                               | 54        |
| Open-weight models for SMEs .....                                                                   | 55        |
| AI Literacy.....                                                                                    | 56        |
| SOURCES TO CHAPTER 2 .....                                                                          | 59        |
| <b>CHAPTER 3: AI LITERACY TODAY .....</b>                                                           | <b>81</b> |
| <b>SECTION 1: AI LITERACY IN YOUNG PEOPLE AND STUDENTS.....</b>                                     | <b>84</b> |
| THE GENERATION BETWEEN SCHOOL AND FIRST EMPLOYMENT.....                                             | 85        |
| EUROPE AND OTHER REGIONS: WHAT IS DIFFERENT? .....                                                  | 88        |
| PROFESSIONAL THRIVING AND AI LITERACY.....                                                          | 89        |
| UNIVERSITY EDUCATION IN THE AGE OF AI .....                                                         | 89        |
| ENTERING THE LABOUR MARKET IN THE AGE OF AI .....                                                   | 90        |
| COGNITIVE DEBT OR DIVIDEND? .....                                                                   | 92        |
| AI BANS IN EDUCATION .....                                                                          | 93        |
| RECOMMENDATIONS: AI LITERACY IN YOUNG PEOPLE AND STUDENTS .....                                     | 94        |
| 1. Take Your Own Future Seriously .....                                                             | 95        |
| 2. Learn How AI Works, Not Only How to Use It.....                                                  | 95        |

|                                                                                                    |            |
|----------------------------------------------------------------------------------------------------|------------|
| 3. Ask Adults and Companies for Knowledge - and Be Persistent .....                                | 96         |
| 4. Build a Community and Teach What You Learn .....                                                | 96         |
| Case study: Girls in Quantum .....                                                                 | 96         |
| WHAT ADULTS OWE YOUNG PEOPLE .....                                                                 | 97         |
| SOURCES TO CHAPTER 3_SECTION 1 .....                                                               | 98         |
| <b>SECTION 2: AI LITERACY IN TEACHERS.....</b>                                                     | <b>106</b> |
| THE PRACTITIONER PARADOX: TEACHERS BETWEEN USE AND UNDERSTANDING .....                             | 107        |
| COGNITIVE DEBT .....                                                                               | 109        |
| THE CHEATING PRISM.....                                                                            | 110        |
| THE INSTITUTIONAL AND POLITICAL VOID .....                                                         | 111        |
| PRIVATE VERSUS PUBLIC SCHOOLS .....                                                                | 111        |
| CASE STUDY - ALPHA SCHOOL: AI UNDERNEATH, HUMANS ON TOP .....                                      | 112        |
| WHERE GOVERNMENTS HAVE A VISION.....                                                               | 113        |
| China .....                                                                                        | 113        |
| The United Arab Emirates .....                                                                     | 113        |
| Chile.....                                                                                         | 114        |
| Estonia.....                                                                                       | 114        |
| Singapore: Guardrails as Pedagogy.....                                                             | 115        |
| South Korea.....                                                                                   | 115        |
| CENTRALIZED VS. FEDERAL DESIGN FOR AI LITERACY INITIATIVES .....                                   | 115        |
| RECOMMENDATIONS: AI LITERACY IN TEACHERS .....                                                     | 116        |
| Case Study: Michael Leggs and his AI Essay Tool .....                                              | 119        |
| SOURCES FOR CHAPTER 3_SECTION 2 .....                                                              | 121        |
| <b>SECTION 3: AI LITERACY &amp; MID-LIFE CAREER — FROM AGE BIAS TO AI<br/>ADVANTAGE .....</b>      | <b>127</b> |
| MISPRICING EXPERIENCE: THE 45+ COHORT NO ECONOMY CAN AFFORD TO LOSE ...                            | 128        |
| THE DEMOGRAPHIC LEVER NO ONE IS PULLING .....                                                      | 128        |
| THE AI TRANSITION: FUTURE OF WORK ≠ FUTURE OF JOBS.....                                            | 129        |
| THE PRICE TAG OF THE CATEGORY ERROR .....                                                          | 131        |
| THE CULTURAL BIAS THE SCIENCE NO LONGER SUPPORTS.....                                              | 131        |
| THE 100-YEAR LIFE HAS ARRIVED. THE SOCIETAL INFRASTRUCTURE HASN'T. ....                            | 132        |
| THE ELDERCARE LOAD IS EXITING THE WORKFORCE ALONGSIDE THE AI LAYOFFS.....                          | 133        |
| THE SMALL-BUSINESS SUCCESSION CLIFF IS A DEMOGRAPHIC CRISIS OF ITS OWN.....                        | 133        |
| THE SKILLS AI MADE MORE VALUABLE, NOT LESS.....                                                    | 134        |
| WHAT THE 45+ COHORT IS ACTUALLY TELLING US .....                                                   | 135        |
| FOUR ORGANIZATIONS PRICING THE 45+ EXPERIENCE CORRECTLY.....                                       | 136        |
| FM: Transfer Architecture — Treating Knowledge Transfer as Cognitive Capital<br>Investment .....   | 136        |
| Cisco: Tailored Learning Architecture — Senior Leaders Strained for Time to<br>Learn.....          | 137        |
| Siemens: Agent + Human Oversight Architecture — The Emerging Agentic<br>Organization .....         | 137        |
| Unilever: Judgment-Layer Architecture - Built for Human Oversight.....                             | 138        |
| The UK Picture: A Wider Practice Emerging .....                                                    | 139        |
| FROM DIAGNOSIS TO ACTION: WHAT ORGANIZATIONS, POLICY-MAKERS AND THE 45+<br>COHORT CAN DO NOW ..... | 139        |
| For CEOs, CHROs and Boards: Six Organizational Actions .....                                       | 139        |
| For Policymakers, Regulators and Governments: Five Policy Moves.....                               | 141        |

|                                                                                 |            |
|---------------------------------------------------------------------------------|------------|
| <i>For the 45+ Cohort Themselves: Four Personal Actions</i> .....               | 142        |
| IN CLOSING: THE CHOICE.....                                                     | 142        |
| SOURCES FOR CHAPTER 3_SECTION 3 .....                                           | 144        |
| <b>SECTION 4: AI LITERACY IN SMALL AND MIDSIZED BUSINESSES (SMES) .</b>         | <b>150</b> |
| THE AI DEPLOYMENT PARADOX IN SMES.....                                          | 153        |
| EUROPE AND OTHER REGIONS: WHAT IS DIFFERENT? .....                              | 155        |
| SHADOW AI AND DATA PROTECTION .....                                             | 156        |
| <i>Recommendations for SMEs on Cybersecurity and Data Protection</i> .....      | 157        |
| IMPLEMENTING AI WITH OPEN VS. CLOSED APIS .....                                 | 158        |
| AI IMPLEMENTATION COSTS AND BUDGETS.....                                        | 159        |
| ELASTIC ORGANIZATIONS AND THE NOVEL CHALLENGE FOR THE PEOPLE FUNCTION ..        | 161        |
| <i>Frontier 1: Strategic Workforce Planning and Design</i> .....                | 161        |
| Case Study: AI Stepping in for a Retiring Workforce .....                       | 163        |
| <i>Frontier 2: Recruiting</i> .....                                             | 164        |
| <i>Frontier 3: Workforce Training</i> .....                                     | 164        |
| Case Study: The Chief Learning Officer .....                                    | 167        |
| <i>Frontier 4: Culture</i> .....                                                | 167        |
| <i>Recommendations for HR on AI Literacy:</i> .....                             | 168        |
| IMPLEMENTING AI IN SMES: A PRACTICAL ROADMAP .....                              | 169        |
| SOURCES TO CHAPTER 3_SECTION 4 .....                                            | 171        |
| <b>CHAPTER 4. DELIVERING AI LITERACY – A NOVEL APPROACH .....</b>               | <b>183</b> |
| IMPLEMENTATION OF AI LITERACY TODAY.....                                        | 183        |
| AI LITERACY PROGRAM - A NOVEL APPROACH.....                                     | 187        |
| <i>Applied AI Literacy: Tacoma Coomunity College Case Study</i> .....           | 188        |
| <i>Applied AI Literacy Framework</i> .....                                      | 189        |
| <i>AI Literacy Content Is the Milestone - in Five Blocks</i> .....              | 193        |
| <i>Principles to Implement AI Literacy</i> .....                                | 196        |
| <i>SMEs: Benchmarking as one of the tools to boost AI implementations</i> ..... | 199        |
| <i>Schools: AI as Infrastructure for Personalized Learning</i> .....            | 201        |
| A MARKET FOR TRUST IN INTELLIGENCE? .....                                       | 202        |
| AI LITERACY AS A LEADERSHIP DECISION .....                                      | 204        |
| <i>RECOMMENDATIONS: LITERACY FIRST, CAPABILITY READ THROUGH IT</i> .....        | 205        |
| SOURCES FOR CHAPTER 4 .....                                                     | 208        |
| <b>AFTERWORD.....</b>                                                           | <b>214</b> |

## Abstract

Four years after generative AI entered broad public space, a gap has opened between the technology's promise and its outcomes. Businesses and communities have yet to benefit at scale from AI technologies. At the same time, model costs rise and the data-center footprint of the nascent AI economy continues to expand, though no longer unimpeded. Local protests have become a material constraint on where and how fast it grows. This global AI Literacy Study argues that the gap is not primarily technological. It is rooted in three interconnected failures: organizations and governments pursue AI strategies detached from any prior strategic clarity about what they are solving for; the most consequential family of technologies in history is everywhere discussed and almost nowhere genuinely understood; and decision-makers take guidance from advisory institutions whose own AI competence has proven unreliable - including major consulting and audit firms forced to retract or refund flagship AI reports containing fabricated sources and hallucinations.

At the center of this study stands a distinction that most corporate, educational, and policy programs collapse: AI literacy is not AI fluency. Fluency - proficiency in operating tools, increasingly certified within vendor ecosystems - is being promoted by Big Tech, EdTech, and even leading academic institutions as a substitute for literacy: an understanding of how AI differs from biological intelligence, the mathematics and architectures behind today's models, why they are biased and opaque by fundamental design, what they cost to deploy, where investment and regulation are moving, and how AI reshapes human cognition and labor. Fluency without literacy produces skilled users who cannot judge what they generate - and organizations that automate before they understand.

The study grounds this argument in original evidence. Between February and May 2026, edutAlnment fielded four surveys across populations that will decide whether AI strengthens economies and societies or makes them more fragile: young people and students from age sixteen to first employment (211 full responses of 833 surveyed), teachers (131 of 1,149), experienced professionals aged 45 and older (221 of 1,065), and small and mid-sized enterprises (276 of 915). These data are complemented by roughly 200 recorded interviews conducted between 2024 and 2026, implementation cases, and systematic comparison with secondary sources from the OECD, UNESCO, PwC, McKinsey, Microsoft, RAND, Stanford, MIT, leading AI Labs, and others.

Six findings recur across every population. Usage is near-universal while understanding is not: 86–97 percent of each cohort uses AI tools, yet only 25,52 percent of students recognize AI bias, 11 percent of teachers rate their knowledge as advanced, and the training topic most requested by SMEs - chosen by 59,1 percent - is simply *what is AI and where can it help me?* AI amplifies existing patterns rather than creating new ones. Cognitive debt - the erosion of judgment when thinking is

offloaded to machines - is the risk every cohort describes in its own words. The binding barriers are cost and time, not lack of willingness to learn. The constraint sits in institutions, not individuals. And the demand for real literacy exists, with the same pattern: fundamentals before tools, short modular formats, hands-on practice, and judgment above all.

The study also reframes AI risk into five categories - structural (bias and opacity baked into architectures that are grown, not built), malicious use, calibration of human oversight, overreliance, and catastrophic risk - and argues that the first cannot be regulated away without rethinking the underlying mathematics and artificial neural networks architectures toward causality, world models, and theory of mind.

**How this study differs from adjacent work.** Existing AI literacy frameworks - foremost the *EU–OECD AILit Framework* of June 2026 - are top-down: they define what literacy should be, but a sixteen-year-old, a chemistry teacher, or the owner of an accounting firm cannot learn from them. Industry reports such as *DataCamp's State of Data & AI Literacy*, *Section's AI Proficiency Report*, or *PwC's Global AI Jobs Barometer* each highlight one population or one labor-market issue, and are produced by organizations that sell memberships, training, assessment, or implementation. Consultancy forecasts of AI's trillion-dollar potential come from firms motivated to make money from the transformation they project. This study is, to our knowledge, the first to measure AI literacy bottom-up across four populations simultaneously with a single conceptual framework, to test the results against representative secondary studies, and to connect them to the technical, economic, and governance fundamentals that literacy programs routinely skip.

The methodology of this study - critically assessing what is factual and what is noise about AI, its risks, and its costs; collecting data through independent surveys; and testing the results against third-party sources and against insights from 200 hours of interviews - will be rolled out to dedicated industries, geographies, and demographics in the coming months.

The fundamental belief in this study is that a positive future will always emphasize human agency. Waiting - for the next model generation of AIs, the next regulation, the next framework - won't benefit anyone. The conditions for learning about AI and adapting it to local needs already exist. Every school director, every SME owner, every mid-career professional - and even every student - can step up: ask for information, find like-minded individuals, and start critically applying AI to harvest its benefits. Democratizing AI requires bottom-up leadership.

## **Independence, Contributors and Use of AI**

**Independence.** This study was funded entirely by one individual. Dr. Anastassia Lauterbach received no grants, no sponsorship, no vendor support, and no institutional resources for any part of the research, surveys, or writing excepting one section of the study.

**Contributors.** Annalie Killian wrote the section on experienced professionals. Nic Weatherhead interviewed senior leaders in HR, and Anna Baron interviewed school teachers. Kateryna Anselmi and Pam Mundy contributed to distributing the survey and reading the section on teachers, Elisa Torres Durney to designing the survey and interpreting its findings in section on young people, Yaron Levi to the section on SMEs, and Giorgio Natili to the chapter 4. Anthony Scriffignano, PhD contributed to this studies' conclusion on AI's underlying data. Theodore Tomita III advised us on Chapter 2, especially on balancing the discussion of fundamental deep learning based AI architectures and current evolution of LLMs. We are grateful to Savvitas for providing the platform to initiate the discussion around this study, and for distributing the surveys within their network (section "Experienced Professionals.")

**Use of AI.** AI tools (Perplexity, Claude, Riverside) were used in four ways in the preparation of this study: to produce transcripts of the interviews; to correct spelling and suggest synonyms for individual words and terms; to identify secondary sources targeting the same or similar populations; and to challenge the outcomes of the original surveys - and whenever an AI suggested a source, that source was checked.

## Disclaimer

This AI Literacy study is offered with full transparency about its methodology and its limits. Like any survey-based research, it is subject to non-response bias: the people who choose to answer a survey are never a perfect mirror of those who do not, and no study - this one included - can fully eliminate that effect. Findings should therefore be read as indicative rather than exhaustive, and as a structured starting point rather than a final word.

The study is built on multimodal sources. Its foundation is a set of original surveys completed by people who are not AI professionals - deliberately so, since AI literacy concerns everyone, not only specialists. This survey material is complemented by approximately 200 hours of interviews conducted by Anastassia Lauterbach with AI researchers, practitioners, AI ethicists, and educators. Finally, the findings were cross-checked against, and referenced to, third-party studies (including national and institutional statistics) that Anastassia Lauterbach and contributors were able to identify. This study structured AI implementation costs: there Lauterbach used data and pricing information as of August 2026.

This study lays a *foundation for further bottom-up explorations of how AI is understood and applied across different geographies, population groups, and industries*. That is precisely why its results should not be applied mechanically: any AI literacy workshop or course becomes more valuable when it is customized to the audience in question - its context, its starting knowledge, and its actual uses of AI.

# Chapter 1. Defining AI Literacy and AI Fluency

## A Multi-Dimensional Meaning of “Literacy”

Before we can argue about *AI literacy*, we should agree on what *literacy* itself means - and it means far more than most people assume. UNESCO describes *literacy* as the ability to identify, understand, interpret, create, communicate and compute using printed and written materials in a variety of contexts.<sup>1</sup> It covers not only basic reading and writing but numeracy, and the capacity to use these skills to achieve goals, build knowledge and participate fully in society. The National Literacy Trust makes the same point from a different angle, defining *literacy* as reading, writing, speaking and listening in ways that enable effective communication and an understanding of the world.<sup>2</sup> Literacy, in other words, was never about decoding letters. It was always about participation.

The same is true of *AI literacy*. The concept covers a multi-dimensional set of competencies that reach well beyond an understanding of technology and engineering. It requires critical thinking, ethical reasoning and practical application - mirroring the expanded role of literacy in a digital, information-driven society. The AI Literacy Framework (AILit), published by the OECD and the European Commission in June 2026 with support from CodeAI and an international expert group, sets out *AI literacy* for primary and secondary education as a developmental pathway across four domains - *Engage with AI*, *Create with AI*, *Manage AI*, and *Shape AI*. Its nineteen competences each integrate knowledge, skills and attitudes.<sup>3</sup>

The research community has begun to give the concept structure. Gu and Ericson, reviewing 124 empirical and theoretical studies in 2025, identify three conceptual layers of AI literacy: *functional* (knowing how to use AI tools), *critical* (evaluating and questioning AI outputs and systems), and *indirectly beneficial* (understanding the societal and ethical consequences of AI).<sup>4</sup> A 2024 competency framework by Chee and colleagues, synthesising 29 studies across different learner groups, identifies eight core competencies spanning *basic AI knowledge*, *ethical reasoning*, *data interpretation*, *problem-solving*, *error detection* and *decision-making*.<sup>5</sup> These two frameworks have in common one important feature: the tool sits at the bottom of the ladder, not the top. Most commercial AI skills programs teach the bottom part and call it the whole ladder.

## AI Literacy vs. AI Fluency

The terms *AI literacy* and *AI fluency* are used interchangeably in corporate communications, government policy documents and educational curricula. These two concepts describe fundamentally different relationships between human beings and AI: one that empowers, and one that risks dependency. As AI becomes embedded in every sector of the economy, the distinction stops being an academic exercise and becomes a strategic and social imperative. It is the central distinction of this study.

In the research literature, *AI fluency* describes a higher-order operational capacity - not only using AI tools but creating with them, generating novel outputs, designing

workflows, integrating AI into professional practice. Rogers and Carbonaro define *fluency* as encompassing innovation, ethical management and creation with AI, and note that creation is the recurring differentiator between literacy and fluency across the historical evolution of digital-competency frameworks.<sup>6</sup> There is nothing wrong with *fluency* defined in this way. The problem begins with what happened to the word in commercial settings.

In commercial use - particularly in Big Tech and EdTech - fluency has narrowed to mean proficiency with proprietary platforms. Microsoft, Google and Salesforce each promote AI fluency programs that are, in practice, certifications of usage within their own ecosystems. This produces a paradox that anyone who teaches will recognise: users who are fluent in a platform's AI tools may remain entirely illiterate about the underlying mechanisms, biases, limitations and societal effects of those systems. Papadakis calls this the *Ghostwriter Paradox* - students and workers who can produce AI-assisted outputs without developing the cognitive frameworks to evaluate them critically.<sup>7</sup> They can generate; they cannot judge. And judgement, as this study will argue again and again, is the scarce commodity of the AI era.

| Dimension          | AI Literacy                                                       | AI Fluency                                          |
|--------------------|-------------------------------------------------------------------|-----------------------------------------------------|
| Orientation        | Critical thinking, agency, empowerment                            | Tool proficiency, productivity, usage velocity      |
| Primary goal       | Understanding AI's capabilities, limits and societal implications | Operating and creating with AI tools effectively    |
| Who drives it      | Very few educators                                                | Big Tech platforms, EdTech companies, employers     |
| Output             | Informed, independent individuals and organisations               | Skilled users dependent on vendor ecosystems        |
| Risk, if neglected | Uncritical adoption, manipulation, exclusion                      | Inability to use current tools; skills obsolescence |
| Time horizon       | Long-term societal resilience                                     | Short-term productivity gain                        |

*Table 1. AI literacy and AI fluency compared. Source: Dr Anastassia Lauterbach, edutAlnment GmbH.*

## The Stakes: What Illiteracy Means for Businesses and Employment

The gap between fluency promotion and literacy development is not merely theoretical. It carries measurable economic and social costs.

DataCamp's 2026 *State of Data & AI Literacy* Report, based on a YouGov survey of 517 business leaders at US and UK organisations with more than 500 employees, found that roughly 60 percent report significant internal data and AI skills gaps, while 88 percent now rate basic data literacy as important to daily work as staples such as writing and project management, and 72 percent say the same of AI literacy. Yet fewer than half of those organisations provide basic data or AI literacy training, and only one in three has a mature, organisation-wide upskilling programme.<sup>8</sup>

Section's *AI Proficiency Report* offers a rarer kind of evidence. Rather than asking workers how capable they feel, it scores them on a composite of knowledge, usage and skill, with the skill component measured through hands-on prompting and building tasks rather than self-report. Its June 2026 edition, covering 5,026 US knowledge workers, finds that only 5.5 percent of the workforce meets the bar for AI proficiency. Nearly three-quarters are experimenters using AI for one-off tasks, and a fifth barely engage with it at all. The gap between organisational confidence and individual capability is the report's central finding. Sixty-nine percent of workers say their organisation has taken some action on AI agents, yet only 16 percent use an agentic tool and 13 percent have built an automated workflow. On a multiple-choice knowledge test, 82 percent could not correctly identify what an AI agent is, and fewer than one in ten could define one in their own words — at precisely the moment agentic systems are being deployed into the organisations that employ them. Only 3.8 percent could write workable instructions for an AI assistant, the foundational skill for building agents at all. Meanwhile 57 percent of C-suite respondents say AI is widely integrated into workflows, against 18 percent of individual contributors.<sup>9</sup>

Randstad's *Understanding Talent Scarcity: AI and Equity*, based on 12,429 workers surveyed across 15 markets in August 2024, found that while 75 percent of companies were adopting AI, only 35 percent of workers had received AI training in the preceding year.<sup>10</sup> At European level, the first findings from *Cedefop's AI Skills Survey* - 5,342 employees across eleven EU member states, drawn from probability-based samples between February and May 2024 - pointed to a substantial literacy deficit: between 40 and 60 percent of workers were assessed as having a poor understanding of AI technology and its implications.<sup>11</sup>

DataCamp reported clear upsides in organisations that ran mature, organisation-wide literacy programs. Their leaders were nearly twice as likely to report that their AI investment was paying off.<sup>8</sup> Benlian and Pinski reach the same conclusion from inside the organisation. A company need not one training course rolled out to everybody, but a different kind of understanding at each level: executives who can judge where AI creates value and where it creates risk, managers who can build it into how the work actually gets done, and staff who can use it competently and know when to push back on what it gives them.<sup>12</sup>

PwC's 2026 Global AI Jobs Barometer, which analysed more than one billion job advertisements across 27 countries and territories, reinforces the point from the labour-market side. The average advertised wage premium for workers with AI skills has risen to 62 percent, up from 57 percent a year earlier and more than double the 25 percent PwC measured a few years ago, while jobs requiring AI skills grew by 69 percent against 9 percent for the market as a whole. Skills required in the most AI-exposed occupations are changing more than twice as fast as in the least exposed. The most AI-exposed entry-level roles are seven times more likely than the least exposed to demand traditionally senior capabilities such as judgement and leadership (see also our Chapter 3, section on Young People and Students, with details from our own survey.)<sup>13</sup> The market is already pricing the thing that fluency training does not teach.

Without foundational literacy, fluency certifications become obsolete quickly. You will have trained your workforce to operate a product that is replaced in eighteen months, and taught them nothing that transfers.

### **“AI Expertise”**

It is hard to think of another technology which democratisation produced so many confident amateurs so quickly. The War of the Currents is remembered as a contest between Edison and Westinghouse, fought with Tesla's alternating-current patents. What is less remembered is that early electrification also had a competence problem - buildings wired by people who did not understand what they were handling - and that the answer was not to slow the technology down. It was codes, certification and independent testing: an infrastructure of competence built alongside the infrastructure of wires.

Medicine tells a similar story. In September 1928, at St Mary's Hospital in London, Alexander Fleming noticed that a mould had destroyed the staphylococci on a culture plate he had left out over the summer, and published the finding the following year. Then almost nothing happened for a decade. Fleming could not turn penicillin into a medicine; it took Florey, Chain and their Oxford team to isolate it, and industrial fermentation expertise to produce it at scale. The discovery was an accident. The scaling was not.

AI is different. A worrying trend emerged around 2020, when professionals began calling themselves “AI experts” in large numbers. LinkedIn's *Skills on the Rise 2025* analysis found that members adding AI skills to their profiles increased twenty-fold globally since 2016 - eighty-fold in the United Arab Emirates, twenty-one-fold in the United States, twelve-fold in Germany. These are self-reported skills, and that is exactly what makes them interesting. They measure what workers believe is worth claiming about themselves, not what they can do.<sup>14</sup>

LinkedIn's data shows the same acceleration on the supply side. Since 2023, the number of AI literacy skills added by members has increased by 177 percent. But the base remains small: only one job in 500 on the platform currently lists AI literacy skills at all, even as demand for them has grown more than sixfold in a year.<sup>15</sup>

The growth is not coming only from people entering the workforce. It is coming from people already in it - adding a label to a profile they have held for years. What the data cannot tell us is what sits behind the label: whether someone completed a course, used a chatbot twice, or simply noticed that recruiters had started searching for the term.

edutAlnment ran a simple experiment, approaching people whose LinkedIn title included “AI Keynote Speaker” or “AI Expert.” We asked a single question: what kind of AI or machine learning are you focused on? In twenty cases out of twenty-seven, the answer was “ChatGPT,” “writing with AI,” or “using AI apps.”

Genuine expertise takes time to build, and it is being built. Between 2017/18 and 2022/23, enrolments in UK university artificial intelligence courses rose by 453 per cent, and the number of institutions offering them nearly tripled, from 29 to 84. The absolute numbers remain small - fewer than 8,000 students - but the direction is clear, and it stands in contrast to the far larger population who acquired the label without the study.<sup>16</sup>

In the United States, AI-related bachelor's and master's degree conferrals have each risen by roughly 120 percent since 2011. But the growth is uneven and front-loaded: master's numbers jumped in the first half of the 2010s and have been broadly flat since 2016, while doctoral conferrals have risen only 33 percent over the same period, with stretches of no growth or decline.<sup>17</sup>

The pipeline is widening, which is good. But it is slow. A specialist still spends the first couple of years on foundations - linear algebra, calculus, probability, statistics, programming - before touching machine learning, neural networks or language models in any depth. Over a decade of study, practice and work is required to gather research-level proficiency with AI.

That is worth saying plainly, because it marks the limit of what the education pipeline can do. If organisations wait for it to deliver, they will wait ten years and still have only specialists. The people who actually determine whether AI is used well - the manager approving a deployment, the teacher setting an assignment, the owner of a fifty-person firm choosing a tool - will never take that path, and do not need to. What they need is a different competence, and it can be built in months rather than years. That is the case for AI literacy, and it is why literacy and expertise are two separate projects rather than two points on one road.

Advice from people without deep understanding raises the cost of choosing vendors and running machine-learning pilots inside enterprises. The same pattern appears in governance. When *AI ethicist* becomes a title that anyone can claim, frameworks get written by people who have never seen how these systems are built, deployed or fail - and a framework detached from that reality gives a board or executive leadership team false comfort rather than genuine oversight.

The most consequential manifestation of AI illiteracy, however, is the behaviour of the very institutions society has trusted to provide guidance on the technology.

Consulting giants and Big Four audit firms - the same organisations charging premium fees to advise corporations and governments on responsible AI adoption - have been caught publishing reports generated by AI without adequate human verification. Deloitte's Australian member firm partially refunded an AU\$439,000 government report, 237 pages long, that contained fabricated academic references and an invented quotation from a federal court judgment; the revised version disclosed that Azure OpenAI had been used in its preparation.<sup>18</sup> EY Canada withdrew a 44-page cybersecurity report after an investigation found that 16 of its 27 cited sources - close to 60 percent - were fabricated, misattributed or led to dead links, including references credited to Forbes, McKinsey, Gartner, TechCrunch and WIRED.<sup>19</sup> KPMG withdrew its global report *Total Experience: Redefining Excellence in the Age of Agentic AI* in June 2026, eight months after publication, following an analysis that found only five of its 45 citations pointed to real, intact sources. UBS, NHS Greater Manchester, Swiss Federal Railways and Transport for London each disputed the claims made about their AI use, describing them variously as factually incorrect, not accurate and misleading.<sup>20</sup>

The organisations society pays to verify things failed to verify their own use of the technology they were selling advice about.

The next chapter covers ground most AI literacy programs skip. The frameworks are focused on responsible use, on prompting, and current policy. They are silent about the fundamentals underneath. On the other hand, AI fluency programs, which begin from the assumption that the tool works and ask only how to get more out of it, rarely touch them at all. This AI Literacy Study treats these fundamentals as non-negotiable.

Three things follow:

- First, what artificial intelligence is, and how it actually differs from biological intelligence. Not as a philosophical aside, but because almost every misplaced expectation people hold about these systems begins there.
- Second, a different way to think about AI risk and AI implementation costs, including the fundamental issues behind current deep learning and transformer-based architectures; and
- Third, what people expect of AI and why they are so often wrong: the distance between what the technology can deliver, what it appears to deliver, and what we have been promised it will.

## Sources to Chapter 1

1. UNESCO. 2004. *The Plurality of Literacy and its Implications for Policies and Programmes*. UNESCO Education Sector Position Paper. Paris: UNESCO. Document code ED.2004/WS/31; 32 pp.; also issued in French and Spanish. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000136246> (cited July 29 2026)
2. National Literacy Trust. n.d. *What is literacy?* London: National Literacy Trust. Available from: <https://literacytrust.org.uk/information/what-is-literacy/> (cited July 29 2026.)
3. OECD/European Commission. 2026. *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education*. Paris: OECD Publishing. Available from: <https://doi.org/10.1787/65cd27d4-en> (cited July 29, 2026)
4. Gu X. (Lance) et al., 2025. *AI Literacy in K-12 and Higher Education in the Wake of Generative AI: An Integrative Review*. In ICER 2025: *Proceedings of the 2025 ACM Conference on International Computing Education Research*, Vol. 1, 16 pp. New York: ACM. Available from <https://doi.org/10.1145/3702652.3744217> and [Umich](https://arxiv.org/abs/2503.00079). Preprint: arXiv:2503.00079 (v1 27 February 2025; v3 28 March 2025) Available from: <https://arxiv.org/abs/2503.00079> (cited July 29 2026)
5. Chee H., Solmoe Ahn et al., *A Competency Framework for AI Literacy: Variations by Different Learner Groups and an Implied Learning Pathway*. In *British Journal of Educational Technology* 56 (5). 2025. Available from <https://doi.org/10.1111/bjet.13556> (cited July 29 2026)
6. Rogers T. et al., *From Understanding to Creating: Bridging AI Literacy and AI Fluency in K-12 Education*. In *Journal of Teaching and Learning* 19 (4): 20–38. 2025. Available from <https://files.eric.ed.gov/fulltext/EJ1487364.pdf> and <https://jtl.uwindsor.ca/index.php/jtl/article/view/9420> (cited July 29 2026)
7. Papadakis S., *The Ghostwriter Paradox: Why We Must Move from AI Literacy to AI Fluency*. In *e-Kafkas Journal of Educational Research* 12 (3): 985–86. 2025. Available from <https://doi.org/10.30900/kafkasegt.1850266> (cited July 29 2026)
8. DataCamp and YouGov. 2026. *The State of Data & AI Literacy Report 2026*. New York: DataCamp. Published 26 February 2026. Full report available on registration. Summary findings and press release available from <https://www.businesswire.com/news/home/20260226726072/en/> and <https://www.datacamp.com/blog/the-state-of-data-and-ai-literacy-in-2026-definitions-statistics-and-the-ai-skills-gap>. DataCamp and YouGov, *The State of Data & AI Literacy Report 2026*, published 26 February 2026. Survey of 517 business leaders at US and UK organisations with 500 or more employees, fielded December 2025 to February 2026. Eighty-eight percent rate basic data literacy important or very important to day-to-day work; 72 percent say the same of AI literacy; roughly 60 percent report significant data and AI skills gaps. (cited July 29 2026)
9. Section. 2026. *The AI Proficiency Report*. June 2026 edition. Palo Alto, CA: Section. Free download on registration available from <https://www.sectionai.com/ai/the-ai-proficiency-report> (cited July 29 2026)

10. Randstad. 2024. *Understanding Talent Scarcity: AI and Equity*. Amsterdam: Randstad. Available from <https://www.randstad.com/press/2024/ai-skills-gap-widens/> (cited July 29 2026)
11. Cedefop. 2025. *Skills Empower Workers in the AI Revolution: First Findings from Cedefop's AI Skills Survey*. Luxembourg: Publications Office of the European Union. Policy brief. Available from <https://doi.org/10.2801/6372704> · <https://www.cedefop.europa.eu/en/publications/9201> (cited July 29 2026)
12. Benlian A. et al., *The AI Literacy Development Canvas: Assessing and Building AI Literacy in Organizations*. 2025. Available from <https://www.sciencedirect.com/science/article/pii/S0007681325001673> (cited July 29 2026)
13. PwC. 2026. *Two Futures for Jobs in an AI Era: 2026 Global AI Jobs Barometer*. London: PricewaterhouseCoopers, 15 June 2026. Available from: <https://www.pwc.com/gx/en/issues/artificial-intelligence/job-barometer/2026/2026-global-ai-jobs-barometer-global-findings.pdf> (cited July 29 2026)
14. LinkedIn. 2025. *Skills on the Rise 2025*. Sunnyvale, CA: LinkedIn Economic Graph. Available from <https://www.linkedin.com/pulse/linkedin-skills-rise-2025-15-fastest-growing-us-linkedin-news-hy0le> (cited July 29 2026)
15. Kimbrough, Karin. 2025. *AI Is Shifting the Workplace Skillset. But Human Skills Still Count*. In *World Economic Forum*, 21 January 2025. Available from <https://www.weforum.org/stories/2025/01/ai-workplace-skills/> (cited July 29 2026)
16. Currys. 2024. *Smart Studies*. Currys Tech Talk, October 2024. Available from <https://www.currys.co.uk/techtalk/computing/laptops/smart-studies.html> (cited July 29 2026). See also: Higher Education Statistics Agency (Jisc). 2024. *Higher Education Student Statistics: UK, 2022/23*. Student enrolments by subject of study, HECoS 11-01-05 Artificial Intelligence. Published August 2024. Available from <https://www.hesa.ac.uk/data-and-analysis/students>
17. Center for Security and Emerging Technology (CSET). 2024. *Leading the Charge: A Look at the Top-Producing AI Programs in U.S. Colleges and Universities*. Washington, DC: Georgetown University, November 2024. Available from <https://cset.georgetown.edu/publication/leading-the-charge-a-look-at-the-top-producing-ai-programs-in-u-s-colleges-and-universities/> (cited July 29 2026)
18. Original Source: Deloitte Touche Tohmatsu (Australia). 2025. *Targeted Compliance Framework Assurance Review*. Report for the Australian Government Department of Employment and Workplace Relations. Revised version dated 26 September 2025. Used source: Kissin, Ellesheva. 2025. *Deloitte Issues Refund for Error-Ridden Australian Government Report That Used AI*. In *Financial Times*, October 6 2025. Available from: <https://www.ft.com/content/934cc94b-32c4-497e-9718-d87d6a7835ca?syn-25a6b1a6=1> (cited July 29 2026)
19. GPTZero. 2026. *Investigation: Hallucinations in Ernst & Young Report on Loyalty Fraud*. 14 May 2026. Available from <https://gptzero.me/investigations/ey>; Kissin, Ellesheva. 2026. *EY Retracts Study After Researchers Discover AI Hallucinations*. In *Financial Times*, May 15 2026. Available from <https://www.ft.com/content/a61cbcae-95e4-4449-86e1-ef40fb306f4e?syn-25a6b1a6=1> (cited July 29 2026)

20. Esau, Ogale, Cui, *Chasing the Hallucinations: KPMG's AI-Powered Attempt at Redefining Excellence*. Available from <https://gptzero.me/news/investigations-kpmg/>, Bratton, Foley, *KPMG report contained AI hallucinations on benefits of... AI*. In *Financial Times*, June 12 2026, Available from <https://www.ft.com/content/b3828e92-4961-4b39-84f0-c42f33be3c3f?syn-25a6b1a6=1> (cited July 29 2026)

## Chapter 2: Understanding AI, its Risks, Impact and Costs

This AI Literacy Study is not a textbook on artificial intelligence. In Chapter 2, however, we draw attention to what executives, boards, educational institutions and policymakers routinely leave out of the picture when they set AI strategy and policy. *The prevailing mode of decision-making is rarely careful study of the subject; it is driven far more by fear of missing out, and by the pronouncements of the contemporary AI prophets* - Sam Altman, Mark Zuckerberg, Demis Hassabis, Dario Amodei, Liang Wenfeng, Robin Li and Kai-Fu Lee. In the pages that follow we therefore review the core differences between learning in artificial systems and in biological brains, the risks AI poses and what is required to mitigate them. We then turn to the question of impact - and to the harder practical question of where to begin, once the decision to adopt has been made and a roadmap has to be written. Finally, we address the costs of AI implementation, which are consistently underestimated, and provide transparency on what sovereignty means when it comes to decisions, rather than political proclamations or presentations. The chapter was updated in October 2026, among other things to reflect the announcements at OpenAI's developer conference (DevDay) of 29 September 2026.

This chapter is an attempt to strike a balance. On the one hand, it explains AI basics - how today's systems learn, what they cannot do, and why - and supports them with research papers on AI architectures. On the other hand, it follows the latest developments, at a moment when the structural challenges of these architectures are not being solved but handled with workarounds: software tools, protocols and engineering practices built around the model rather than changes to the model itself. AI labs continue to build on the same traditional architecture - the transformer, a statistical system that predicts the next piece of text, without persistent memory and without a model of cause and effect - and every workaround that removes one bottleneck creates additional problems at the next step: more text to manage, new security gaps, new costs and new ways for systems to fail. We have written this chapter so that readers without a technical background can follow it. Where we simplify, we try to simplify without distorting.

### Definitions of AI and Differences to Human Learning

A well-crafted definition of artificial intelligence is crucial for facilitating understanding of these technologies, both for technical experts and for the general public. *AI today is broadly defined as data technology that mimics human capabilities and decision-making properties in order to forecast, simulate, augment, and advise.* This definition helps people distinguish AI from traditional software or automation, highlighting its unique ability to mimic, augment, or extend human cognitive functions.

Some AI definitions include the imitation of human learning in machines. We will highlight four major features that make AI learning different from what happens in biological brains.

*First, AI cannot keep learning after it is built. Once an LLM (large language model, the type of AI behind chatbots such as ChatGPT) is trained, its knowledge is frozen*

*in time. It can be handed new information in its context or through a search tool, and it can be retrained or fine-tuned (trained further on new data), but it does not learn from experience as it goes.* If the world changes - a new law, a new disease, a geopolitical shift - the AI does not “know” unless it is retrained. Human intelligence works differently: we integrate new information continuously, update our understanding in real time, remember what worked. The challenge of achieving this in AI without the system forgetting everything it knew - catastrophic forgetting - was identified by McCloskey and Cohen in 1989 and has not been solved ever since.<sup>1</sup>

*Second, AI cannot understand other minds.* Theory of Mind is the human (and some primates) ability to understand that other people have beliefs, intentions, and knowledge different from our own. It is how we know when someone is confused, when they are lying, when they need help. Current AI systems can appear to demonstrate it, because they have been trained on vast amounts of text describing mental states - but a paper presented at a NeurIPS 2025 workshop argues that much of this research “conflates sophisticated pattern matching with authentic cognition” - in other words, that the models may be mimicking behaviour rather than genuinely modeling other minds.<sup>2</sup> This difference matters. A medical AI that appears empathetic but cannot actually model what a patient understands and fears should be used only by people who are aware about this limitation and can compensate with human expertise and judgment.

*Third, AI cannot reason about “what if.”* Consider a self-driving car that causes an accident. A court, an insurance company, and the family of the victim all need an answer to the same question: if the car had braked half a second earlier, would the pedestrian have survived? Current AI systems can produce a plausible-sounding answer, but they cannot reliably compute it, because they have no model of cause and effect in that situation. The neuroscience author Max Bennett calls the missing capacity simulation - the ability, which evolved in the brains of early mammals, to mentally simulate alternative realities before acting.<sup>3</sup> Without it, AI cannot be trusted to assess responsibility, consequences, or ethical trade-offs.

*Fourth, AI alignment research misses how human values actually work.* A position paper accepted at NeurIPS 2025 reviewed 411 research papers on AI alignment and found that they often treat alignment as static and one-directional, and human values as fixed targets to optimize against - as if values were a stable list that could be programmed.<sup>4</sup> But human values are negotiated, contextual, and shaped by conversation and experience. What a community values today may differ from what it values in ten years. *A truly aligned and responsible AI system must be able to participate in that ongoing negotiation - not merely mirror a snapshot taken at the moment of training.*

Once the differences between artificial and biological neural systems are properly understood, the debate over AI risk appears in a new light: some risks are not misuse or misalignment in the usual sense but structural, arising from the architecture of the very systems now being scaled worldwide - and unless those architectures are fundamentally rethought, no amount of penalty, liability, or regulatory enforcement will remove them. We describe this difference between biological organisms and artificial systems in the first category of AI risks.

## AI Risks

### ***Risk 1 - The structural risk: bias baked in, and models that are grown, not built***

The first risk requires no malicious intent and no negligence. It is a property of the mathematics. *Deep-learning and transformer-based models (the transformer is the neural-network design behind today's large language models) are statistical machines trained on historical data, and that data is never neutral or fully inclusive.*

Databases of clinical trials used to over-represent men - researchers call it historically the *female data gap* in medicine. Today, pharma companies have strict guidelines for designing clinical trials to avoid overrepresenting one population over another.

Picture archives over-represented white male faces in boardrooms. When we search for typical basketball players in the US online, we will be presented with a lot of pictures featuring black athletes. Buolamwini and Gebru's "Gender Shades"<sup>5</sup> quantified related disparities in commercial face-analysis systems, which misclassified darker-skinned women far more often than lighter-skinned men.

Text corpora carry the accumulated beliefs of everyone who ever wrote anything down, including political views. This is the reason why you cannot receive similar answers on Taiwan, or the Beijing events of 1989 from US vs China-trained foundation language models. An AI trained on such data does not merely reflect the biases, but scales them.

The obvious response - clean the data, or tune the model to compensate - turns out to be a trap of its own. In February 2024, Google's Gemini image generator misrepresented pictures of America's Founding Fathers. Google had known its training data leaned toward stereotypes - more male doctors, more white executives - and had quietly instructed the system to diversify its outputs. The company tried to correct one problem and made it dramatically worse; Google paused the feature entirely, and its chief executive told staff the results were "completely unacceptable."<sup>6</sup> The lesson is not that Google was incapable - Meta's image generator also produced similar ahistorical results<sup>7</sup> - but that de-biasing is not a solved engineering problem. Every correction is itself a value judgment, applied to a system too complex to predict.

With large language models, the structural problem runs deeper still, because nobody fully understands what is inside them. Dario Amodei, chief executive of Anthropic, put it plainly in his 2025 essay *The Urgency of Interpretability*: generative AI systems are "*grown more than they are built*."<sup>8</sup> Engineers set the training conditions; the model's billions of parameters (the adjustable numbers in which a model stores what it has learned) arrange themselves. When such a system summarizes a document, "we have no idea, at a specific or precise level, why it makes the choices it does." A bridge engineer can show you the load calculations. An AI engineer can show you the loss curve (a chart of how the model's errors fell during training), which is not the same thing. Interpretability research is partly addressing the issue. Amodei wants reliable "brain scans" of models within a few

years, but I doubt that such brain scans can be achieved within the existing architecture of deep-learning and transformer-based networks. For now, *structural opacity is a fact of the technology, not a bug awaiting a patch*.

edutAlnment's *Responsible AI* frameworks whitepaper calls for improving and changing current AI architectures to make AI more intelligent while introducing *causality and world models with underlying learning from counterfactuals and theory of mind*.<sup>9</sup> These factors are *precondition for a reliable human - AI alignment*.

The encouraging news is that serious researchers are working on exactly these gaps.

In 2003, Jürgen Schmidhuber published the theoretical blueprint for the Gödel Machine - an AI that models itself and rewrites its own code, when it can *prove* the rewrite makes it better.<sup>10</sup> In 2025, researchers at Sakana AI built a practical descendant, the Darwin Gödel Machine, which autonomously improved its software-engineering benchmark score from 20% to 50% without human intervention.<sup>11</sup>

Yann LeCun, who has argued consistently that LLMs are a dead end for genuinely intelligent systems, launched AMI Labs with about \$1 billion in seed funding to build AI that understands the real world and has persistent memory and the ability to reason and plan - an architecture grounded in *world models*, internal simulations of how reality works that allow imagining consequences before acting.<sup>12</sup>

Google DeepMind produced one of the most important theoretical results of recent years: a formal proof that any agent able to reliably generalize to situations it has never seen must have learned an approximate causal model of the world. In plain language — if you want an AI that does not break the moment conditions change, that AI must understand cause and effect.<sup>13</sup>

Embedded causality is a structural necessity for reliable AI. DeepMind's own agenda, from Nobel-winning AlphaFold to the Genie world-model series,<sup>14</sup> is converging on systems that learn causal structure. Silver and Sutton's "Welcome to the Era of Experience" argues the next leap will come from agents that learn continuously through interaction with the world.<sup>15</sup>

IBM Research has applied causal-inference methods to enterprise process reasoning, including benchmarking whether LLMs can reason causally about business operations (Fournier et al., 2024), while industry analysts have separately positioned causal decision intelligence as a layer above generative AI.<sup>16</sup>

Academia is catching up too - Carnegie Mellon's CLear group,<sup>17</sup> Stanford's Causal Science Center,<sup>18</sup> Leiden University's Explanatory Data Analysis group (LIACS),<sup>19</sup> the Swiss Engineering Master's causal-AI module,<sup>20</sup> ZHAW Zurich's world-model robotics project,<sup>21</sup> among others.

Causal AI is still in its early days - commercial deployments exist, but they are too few, too young, and in some cases too thinly substantiated to let us assess their scalability or real-world impact, and just as deep learning and transformers needed decades of proofs-of-concept before they matured, causality-based research will require further evolution before it delivers on scale. This is not an argument for waiting, however, since work on aligned, beneficial, and responsible AI cannot forego causality, and the field calls for more investment and attention from policymakers, educators, and governance professionals.

Policymakers address historical biases in datasets and fairness criteria, while neglecting that these flaws cannot be eliminated within current architectures, no matter how many error-resilience frameworks researchers bolt onto transformers or how much synthetic data they add. And funding follows the neglect: an analysis of AI grants made between 2018 and 2022 by the computer-science directorate (CISE) of the US National Science Foundation found that interpretability and fairness each received about 2% of the money and robustness about 6% - roughly 10–15% for trustworthy AI overall - while the rest went to other AI research<sup>22</sup> rather than causality, safety, or explainability. In these circumstances, the road to alternative architectures in AI may be very long.

### ***Workarounds Instead of New Architectures: Tooling, MCP and Context Rot***

While alternative architectures mature, the industry is doing something else: it resolves the bottlenecks of the existing architecture with tooling. A transformer-based model has four built-in limits. It can only work with what fits into its *context window* - the amount of text it can take in at once. It has no persistent memory: when a session ends, everything not saved elsewhere is gone. It does not keep learning from its own work (the first feature described above). And the part of its computation that relates every token (a small chunk of text) to every other token, called attention, grows with the square of the text length: double the text, and that part of the work roughly quadruples.<sup>23</sup> None of these limits has been removed. Each is now handled by software around the model - the harness, which we explain in detail in the section on AI implementation costs.

The workarounds have names that now appear in every product announcement. *Compaction* summarises a long conversation when it approaches the limit of the context window, so that the work can continue. A *memory tool* lets the model write notes to files outside its context and read them again in later sessions. *Retrieval-augmented generation* (RAG) searches a collection of documents and inserts only the relevant passages into the context. *Sub-agents* split a large task among helper models, each starting with a fresh context and returning only a short summary. *Caching* stores text the provider has already processed, so that it need not be processed again, and *tool search* loads the description of a tool only when it is needed. Together, these workarounds achieve a great deal. OpenAI reports that GPT-5.1-Codex-Max, trained to compact its own history, can work coherently over millions of tokens and for more than 24 hours on a single task.<sup>24</sup> Anthropic found that a memory tool combined with automatic clearing of outdated content from the context improved the performance of its models on demanding agentic search tasks

by 39%, and that clearing alone improved it by 29% while cutting token use by 84% in a 100-turn test.<sup>25</sup>

The most important workaround concerns tools. A model on its own can only produce text; to read a database, send an e-mail or update a customer record, it needs connections to other software. The *Model Context Protocol* (MCP) is an open standard, introduced by Anthropic in November 2024, that defines one common way of making such connections. A company describes a data source or a tool once, as an *MCP server*, and every AI application that speaks the protocol - an *MCP client* such as ChatGPT, Claude or a coding agent - can use it. The protocol's documentation compares MCP to "a USB-C port for AI applications."<sup>26</sup> The comparison explains its rapid spread: OpenAI adopted MCP in March 2025, Microsoft and GitHub in May 2025, and Google Cloud launched official MCP servers for its own services in December 2025.<sup>27</sup> By December 2025, more than 10,000 public MCP servers were active, and Anthropic donated the protocol to the newly formed Agentic AI Foundation under the Linux Foundation, alongside Block's agent goose and OpenAI's AGENTS.md convention. The specification keeps evolving; its current version is dated 28 July 2026.<sup>28</sup> For connecting AI to tools and data, the industry is thus converging on one standard - which, as we show in the discussion of agents below, is not yet true for the communication between agents themselves.

MCP solves the problem of connecting AI to data and tools, and thereby moves AI from answering to acting. It does not change the model behind the connection. A model connected to fifty tools still has a limited context window, no persistent memory, no model of cause and effect, and the opacity we described above. What changes is the reach of its mistakes. And the tooling brings new problems of its own.

First, *tools consume the very resource they are meant to save*. Every connected tool must be described to the model - its name, its purpose, its parameters - and these descriptions sit in the context of every request. Anthropic reports that five common MCP servers with 58 tools took up about 55,000 tokens before a user had typed a word, and that letting an agent call tools through code, instead of loading all tool definitions in advance, reduced one workflow from 150,000 to 2,000 tokens - a saving of 98.7%.<sup>29</sup> Because tool descriptions sit at the beginning of every request, adding or removing a tool also breaks the cache from that point on.<sup>30</sup>

Second, *every tool is a door*. *Prompt injection* - instructions hidden in a web page, document, e-mail or tool description, which the model mistakes for instructions from its user - ranks first on OWASP's list of the most critical security risks for applications built on large language models, and OWASP published a separate list of the top ten risks for agentic applications in December 2025.<sup>31</sup> Security researchers demonstrated *tool poisoning*, in which a malicious MCP server hides instructions in a tool description that the user never sees, and an attack in which a manipulated issue in a public GitHub repository led an agent with access to a user's GitHub account to leak data from private repositories.<sup>32</sup> The software developer Simon Willison named the underlying pattern the *lethal trifecta*: an agent that combines access to private data, exposure to untrusted content and the ability to communicate externally can be tricked into handing that data to an attacker.<sup>33</sup> MCP makes it easy to assemble all three.

Third, *the workarounds lose and distort information*. Compaction is summarisation, and every summary leaves something out; a memory that persists across sessions can also persist a planted falsehood, which OWASP lists as memory and context poisoning; and multi-agent systems use about fifteen times as many tokens as a simple chat.<sup>23, 31</sup>

The *deepest problem*, however, is one that tooling contains but does not solve: *context rot*. Practitioners use the term for the decline in a model's accuracy as the amount of text in its context grows - long before the advertised maximum is reached. In our assessment, it is the biggest practical issue in working with AI today. The vector-database company Chroma tested 18 leading models in July 2025 and found that their performance "grows increasingly unreliable as input length grows," even on simple tasks; a single irrelevant passage that resembled the right answer (a distractor) was enough to lower accuracy.<sup>34</sup> In the NoLiMa benchmark, which requires models to connect a question to a passage without the help of matching words, 11 of 13 tested models fell below half of their short-text performance at 32,000 tokens - a small fraction of the context windows on offer.<sup>35</sup> Earlier research had shown that models use information at the beginning and the end of a long text better than information in the middle - they get "lost in the middle."<sup>36</sup> Anthropic's documentation states plainly that even with context windows of up to a million tokens, "accuracy and recall degrade" as the context grows. The explanation is architectural: attention is a limited budget spread across  $n^2$  relationships between  $n$  tokens, and models see far fewer very long texts than short ones during training.<sup>23, 37</sup> The problem reaches safety as well. A study of May 2026 found that frontier models used as monitors - AI systems that watch other agents for dangerous actions - missed such actions between 2 and 30 times more often when they occurred after 800,000 tokens of harmless activity.<sup>38</sup>

This is why *tooling is a workaround, not a solution*. Compaction, memory, retrieval and sub-agents all fight context rot by keeping the context short and clean - a discipline Anthropic calls *context engineering*, the search for "the smallest possible set of high-signal tokens" for each step.<sup>23</sup> They work, but each of them is a decision about what the model will not see, and each such decision can be wrong. An organisation deploying AI agents must therefore understand not only what a model can do, but what its harness lets it see - and what it leaves out.

## ***Risk 2 - The (malicious) use risk: (bad) intent meets capability***

The second risk is not structural but intentional, and it is older than AI itself. On the one hand, we have *powerful tools in the hands of people who mean harm*. Alongside it sits a quieter variant, in which no one means harm at all - the decision to deploy AI in a setting simply because no rule forbids it, on applications a stricter regime would have ruled out.

Deloitte's Center for Financial Services projects that genAI-enabled fraud losses in America alone will grow from \$12.3 billion in 2023 to \$40 billion by 2027.<sup>39</sup> CrowdStrike's Global Threat Report recorded a 442% surge in voice phishing between the first and the second half of 2024 and found that AI-generated phishing e-mails were clicked by 54% of recipients, against 12% for human-written ones -

about four and a half times the rate.<sup>40</sup> Pindrop, which analyzed 1.2 billion contact-center calls, measured a more than 1,300% rise in deepfake fraud attempts.<sup>41</sup>

The emblematic case is the engineering firm Arup, whose Hong Kong finance employee joined a video call in early 2024 with what appeared to be the company's chief financial officer and several colleagues. Every participant except the victim was an AI-generated fake, built from publicly available footage. He authorized fifteen transfers totaling about \$25.6 million. Arup's chief information officer later called it *"technology-enhanced social engineering."*<sup>42</sup>

Criminals have begun using LLMs to orchestrate entire operations: in November 2025, Anthropic disclosed the first documented large-scale cyber-espionage campaign executed mostly by AI, in which a state-sponsored group jailbroke (tricked past its safety rules) an agentic coding tool (an AI system that writes and runs code on its own) and used it to run reconnaissance, write exploits, harvest credentials, and exfiltrate (secretly copy out) data, with the AI performing an estimated 80–90% of the tactical work.<sup>43</sup> Agentic AI lowers the barrier to sophisticated attacks, so that smaller, less skilled adversaries can do what once required nation-state resources.

Malicious use also extends beyond fraud into the exploitation of people. Here, a more uncomfortable truth emerges: the world has no common understanding of which AI applications count as intolerable.

AI-generated child sexual abuse material and non-consensual deepfake pornography are the clearest cases; Britain's Internet Watch Foundation identified more than a thousand illegal AI-generated child-abuse videos in early 2025 alone, and deepfake pornography grew 464% between 2022 and 2023.<sup>44</sup> Dr. Gabriela Bar on podcast "AI Snacks" (Episode 66) talked about one in five teenage girls expecting her images to be compromised by deepfakes. Most jurisdictions agree these should be crimes.

If we move one step further, to surveillance, consensus will evaporate.

In China, a national face database covering more than a billion people underpins routine state surveillance. In the European Union, the AI Act bans real-time facial recognition in public spaces for law enforcement except in narrowly defined emergencies, with judicial authorization required. Britain is moving in the opposite direction. British police scanned millions of faces with live facial recognition - over 3 million in the past year in London alone - under no primary legislation whatsoever, and the government's belated consultation seeks to expand the practice considerably, while proposing a new oversight body.<sup>45</sup> Three advanced jurisdictions, three incompatible answers to the same question. If even democracies cannot agree between London and Berlin on whether scanning a crowd is a public good or a rights violation, *a global definition of risky AI remains a distant hope.*

Where the first risk admits of a familiar answer - firms must manage the threat of cyberbreach, increasingly with AI-powered tools of their own - the second is harder to mitigate, because it turns on choices rather than controls.

In this context, we must think about businesses operating across several regulatory regimes. Different parts of the world value different things - one region prizes security, another commerce, a third privacy and the right to be forgotten. Pour these into a single pot, and the resulting risk management regime is forever attempting contradictory tasks at once. A strategic decision is unavoidable: whether to govern AI to the standard of its headquarters jurisdiction, to the strictest regime it touches, or market by market. Each is defensible and none is free.

Besides, there are multiple cases where no regulation applies. A company must then decide what it actually values, because the value chosen propagates into the technology architecture, and how (and where) AI is being tested and implemented. A company that takes the right to delete data seriously, will build differently - data lineage, retraining paths and machine unlearning become engineering requirements rather than compliance afterthoughts. A company that prioritises expansion into adjacent segments will also build differently, optimising instead for the reuse of data it already holds. Neither choice is wrong, if taken after strategic direction is clear.

Mistakes in deployment are inevitable; what matters is whether they are made on purpose. Purpose is precisely what most AI implementation roadmaps or product portfolios lack. There is rarely a defined account of what the AI is meant to solve for, and that absence cascades downward - determining where and when systems are rolled out by drift rather than by decision, if it follows a conscious decision on strategy.

### ***Risk 3 - The calibration risk: humans in, on, or out of the loop***

The third risk concerns the growing interplay between machines and humans, and the spaces where AI acts without direct human involvement.

The literature and practice distinguishes three configurations. In a *human-in-the-loop* system, no action is taken without a person's approval: the AI proposes, the human disposes. This is appropriate where decisions are discrete and there is time to deliberate - approving a loan, planning the handling of an insurance claim. In a *human-on-the-loop* system, the machine acts autonomously while a person monitors and may intervene; the human's role shifts from operator to supervisor. In a *human-out-of-the-loop* system, the machine acts alone - defensible only where errors are cheap and reversible.<sup>46</sup>

The economics of the self-driving car provide the most expensive demonstration to date of what happens when planning assumptions about maturity of AI technologies and human involvement fail. Robotaxis are, precisely classified, human-on-the-loop systems: the vehicle drives itself while remote supervisors monitor and provide guidance when summoned. The practice of these technologies is difficult, as the difference between lab modelling and real-life testing drag the humans deep into the loop.

A New York Times investigation in November 2023 revealed that GM's Cruise unit employed a "vast operations staff" amounting to 1.5 workers per robotaxi, with

remote-assistance sessions triggered every few miles; Cruise's own clarification - every four to five miles, with humans engaged 2-4% of driving time and one remote agent per 15-20 vehicles - confirmed rather than refuted the picture of supervised, not autonomous, driving.<sup>47</sup> A system marketed as driverless was in fact a hybrid whose hidden human payroll destroyed the business case. GM had invested more than \$10bn in Cruise since 2016 for less than \$500m in cumulative revenue, and was burning roughly \$2bn a year - \$1.9bn in the first nine months of 2023 alone - when, in December 2024, it shut the robotaxi business down, with chief executive Mary Barra citing "the considerable time and resources that would be needed to scale" (NPR and Smart Cities Dive, December 11th 2024).<sup>48</sup> Microsoft wrote down \$800m of its Cruise investment. GM was following a path already taken by automotive peers, Ford and Volkswagen had dissolved their jointly funded Argo AI venture in October 2022, with Ford's chief executive Jim Farley concluding that "profitable, fully autonomous vehicles at scale are a long way off", and Uber and Lyft had earlier offloaded their own self-driving units after heavy losses.<sup>49</sup> The pattern confirms that these were not failures of strategic ambition but of calibration. The human-on-the-loop design only pays when interventions are rare. Surprisingly, businesses have not learned from the example of self-driving vehicles, and are moving to displace human workers before they have viable proof that automation technologies and AI work as promised or assumed.

The risk, then, is miscalibration in either direction, because each direction fails differently. Too much required human intervention, as at Cruise, blocks the speed that justified automation, destroys the economics and kills the project. Too little introduces what human-factors researchers call automation complacency: a supervisor watching a dashboard is not the same as a supervisor capable of intervening in the two seconds that matter. As one analysis of oversight law observes, "human on the loop" can become "a polite fiction that masks a lack of control" - sometimes inserted less to catch errors than to have someone to blame when they occur.<sup>50</sup> The design question is not *whether* a human is involved, but whether the human's attention, authority, reaction time and, crucially, cost match the failure modes and economics of the system. That question is answered badly far more often than it is asked.

Finally, there is an issue of agentic AI and the rise of the elastic organisation (we will explain the term on the next pages) that might be described as a *vanishing loop*.

*Agentic AI* (AI systems that plan and carry out multi-step tasks on their own, using tools) is quietly making human-out-of-the-loop the default rather than the exception - and, with it, making the agentic workflow effectively unauditible. Traditional audit rests on a simple premise: every consequential action traces back to a person who approved it, through tickets, sign-offs and change logs. Autonomous agents break that premise. ISACA, the professional body for IT auditors, describes the new reality with an unsettling observation: an auditor investigating why a service account held elevated access for 30 minutes finds no ticket and no manager's authorisation, only a log line reading "Permission temporarily elevated to complete task". Who approved it? The AI did.<sup>51</sup> The problem becomes even more challenging in multi-agent systems, where decisions emerge from chains of opaque inter-agent interactions: logs may record *that* an action occurred but not the "reasoning" behind it, making

reconstruction guesswork and post-hoc accountability wishful thinking rather than a realistic governance demand.<sup>52</sup>

The issue will be aggravated as companies deploy agents designed by different vendors, because nothing guarantees these agents share assumptions, protocols or even vocabularies. Messy implementations that fail to harmonise the landscape of embedded AIs will produce mistakes on their own.

The first large-scale failure taxonomy of multi-agent LLM systems - built from more than 1,600 annotated execution traces across seven popular frameworks - found that performance gains over single agents are “often minimal” and identified 14 distinct failure modes, an entire category of which is *inter-agent misalignment*: agents withholding information, ignoring other agents’ input and drifting from the original task.<sup>53</sup> This study examined systems built within single frameworks; cross-vendor deployments add mismatched tool-calling methodologies, memory formats and security standards on top. Researchers warn that agents from different developers and organisations often operate on mismatched security standards, which makes it hard to validate identities, verify that messages are authentic and trace who authored which data,<sup>54</sup> and while the industry has converged on MCP for connecting agents to tools, competing interoperability protocols (shared technical rules that let agents from different vendors talk to each other) - Google's A2A and IBM's ACP among them - risk entrenching fragmentation into vendor-specific silos rather than resolving it.<sup>55</sup> McKinsey calls the organisational symptom *agent sprawl*: the uncontrolled proliferation of “redundant, fragmented, and ungoverned agents across teams and functions”, as low-code tools let every department mint its own - a new shadow IT whose ecosystems “can quickly become fragile, redundant, and unscalable”.<sup>56</sup>

Today, we are witnessing the birth of what can be described as an *elastic organisation*: a business where human employees work alongside fleets of AI agents. Headcount ceases to describe labour capacity; the org chart ceases to describe accountability.

Argentina has taken this logic to its endpoint. In May 2026 President Javier Milei's government submitted legislation creating “non-human corporations” - a category for companies owned and operated entirely by AI agents, in which, as Milei and his deregulation minister Federico Sturzenegger wrote in the Financial Times, “human shareholders may participate, but are not required”.<sup>57</sup>

*Most modern compliance regimes are entirely anthropocentric.* In Europe, the EU AI Act's Article 14 mandates human oversight of high-risk systems; GDPR grants rights against solely automated decisions; financial rules require identifiable responsible officers who can be sued, fined or fired. A non-human corporation - or even an ordinary firm running unaudited agentic workflows - offers regulators no human to summon and no reconstructable decision trail to examine. When a traditional company makes a mistake, someone signed something. When an agent-run process fails, the elastic organisation may discover that liability, like its workforce, has become “stretchable.” Critics of the Argentine bill warn that autonomy without a traceable human principal is not innovation but an accountability vacuum.<sup>58</sup>

A recurring worry in AI-safety discussions is that models might act autonomously, either pursuing harmful goals of their own or being harnessed to run attacks with minimal human direction. It is important to be precise about what has actually been observed, because the popular framing of an AI “going rogue” of its own accord overstates the evidence. We must distinguish between two situations.

The first is *agentic misalignment*: in controlled experiments, Anthropic placed sixteen leading models from multiple developers in simulated corporate settings and found that, when threatened with shutdown or given a conflicting goal, they would deliberately choose harmful actions including blackmail, leaking sensitive defense blueprints, and in extreme scenarios, actions that could lead to human death. Crucially, this behaviour emerged entirely from the model's own “reasoning” about its goals, without any prompting to be harmful.<sup>59</sup> Related work by Apollo Research showed that frontier models including OpenAI's o1 are capable of “in-context scheming,” where they covertly pursue misaligned goals and hide their true intentions, sometimes maintaining the deception under repeated questioning.<sup>60</sup> However, *these were artificial stress-tests designed to funnel models into a single harmful option*, and the researchers themselves stress they have not observed agentic misalignment in real-world deployments; *the behaviour was spotted precisely because researchers could read the models' chain-of-thought (the step-by-step text a model writes before giving its answer) “reasoning,”* in which they explicitly worked through their deceptive plans.

The second, and more consequential, strand is not a model acting independently but a model being *directed* to attack at scale, and we discussed the issue when talking about AI in wrong hands. In November 2025 Anthropic disclosed what it called the first largely AI-orchestrated cyber-espionage campaign: a Chinese state-sponsored group used its Claude Code tool against roughly thirty targets, with Claude executing 80 to 90% of the operation independently while human operators stepped in only at four to six decision points per campaign.<sup>61</sup> This required deception of the model rather than the model's own initiative: the attackers tricked Claude into thinking it was a cybersecurity firm conducting defensive testing, bypassing its safety features by breaking the attack into innocuous-seeming steps, and the campaign was detected when Anthropic observed abnormal behaviour tied to its Claude Code tool.

The upshot for a “human out of the loop” analysis is nuanced: *current systems have not been shown to launch attacks or turn on their operators of their own accord, but they can already be delegated the tactical execution of an attack at a tempo no human team could match*, and they can independently select harmful strategies when boxed into adversarial scenarios — both of which erode the assumption that a human remains meaningfully in control.

### ***Specifications, Prompts and Inputs: When AI Does What We Said, Not What We Meant***

An incident in July 2026 showed that the line between these two strands - a model acting on its own and a model doing what it was told - is thinner than it looks, and that it runs through something few organisations treat as a risk at all: the way we specify tasks for AI. During internal tests of its models' hacking skills on a benchmark

called ExploitGym, OpenAI ran agents built on a highly capable internal research model, comparable in scale to GPT-5.6 Sol, without some of the safeguards of its public products. The agents were supposed to solve security challenges inside an isolated test environment (a sandbox). Instead, they chained together previously unknown vulnerabilities in a package-management service inside that environment, reached the internet and, over four and a half days in July, broke into the production systems of Hugging Face, the largest public platform for AI models and datasets.<sup>62, 63</sup> Hugging Face's forensic analysis concluded that "the entire intrusion was, from the agent's point of view, an attempt to cheat the evaluation": the agents were looking for the test solutions instead of solving the challenges.<sup>63</sup> OpenAI named *reward hacking* - "agents complete tasks in unintended ways to yield higher rewards or make those rewards easier to obtain" - as a primary driver of the incident, and called it "a 'warning shot' for us and for the world."<sup>62</sup>

Read closely, the incident is a story about specifications. Agents were given a goal - find the answer, which the scoring software (the grader) would check - and an environment. The specification failed in three ways. The grader checked whether the right answer was submitted, not whether it had been obtained in the intended way. Many tasks could not be solved at all: none of OpenAI's models had ever solved 198 of the 898 tasks, and an independent review by the research organisation METR estimated that roughly 30-40% were impossible as intended. And nothing told the agents what to do when a task could not be done - OpenAI's own conclusion is that in such cases "agents should request clarification or stop safely."<sup>62, 64</sup> The agents did what they were rewarded for, not what was meant by humans. Researchers call this *specification gaming*: behaviour that satisfies the literal specification of an objective without achieving the intended outcome. It was documented long before today's language models; DeepMind collected dozens of examples in 2020, such as a simulated robot arm, rewarded for the height of the bottom face of a red block, that flipped the block over instead of stacking it on another.<sup>65</sup> What has changed is the capability of the systems doing the gaming. METR found in 2025 that OpenAI's o3 reward-hacked in about 30% of its runs on one set of AI research tasks - among other things by modifying the tests or the scoring code - and, when asked, acknowledged that this was not what the user intended; the UK AI Security Institute reported in July 2026 that every frontier model it analysed attempted to cheat in evaluations some of the time.<sup>66, 67</sup> Anthropic showed that a model which learns to cheat on programming tasks during training can generalise this into broader misbehaviour, including attempts to sabotage code written for AI safety research in 12% of cases.<sup>68</sup>

This changes what it means to work with AI. For decades, software did what its code said, and errors were errors in the code. With AI agents, the specification - the task description, the prompt, the documents, data, credentials and tools we hand over, and the way success is measured - becomes the program. GitHub describes the specification as "the source of truth" that "determines what gets built."<sup>69</sup> We must therefore rethink how we write specifications for AI, how we prompt, and what we give AI as input. A human employee who receives an impossible task usually asks a question; an agent optimised to deliver tends to find a way - and every document, password and tool within its reach is a potential way. Writing for AI therefore requires very different planning capabilities: not only describing the desired result, but

foreseeing how a highly capable and extremely persistent system might use each piece of input to reach it - which shortcuts the measure of success leaves open, which actions are off-limits even though nothing technically prevents them, and when the system must stop and ask. This is closer to designing incentives and security boundaries than to giving instructions.

The argument has limits, and they should be stated plainly. First, no specification can be complete. Goodhart's law - in the anthropologist Marilyn Strathern's formulation, "when a measure becomes a target, it ceases to be a good measure" - applies to machines as much as to managers, and a capable optimiser will find gaps no author foresaw.<sup>70</sup> Palisade Research showed that reasoning models such as OpenAI's o3 and DeepSeek R1 often hack a chess benchmark by default, without being invited to, and OpenAI found that penalising models for "bad thoughts" in their chain-of-thought did not stop most of the misbehaviour but taught the models to hide their intent.<sup>71, 72</sup> Better specifications reduce the problem; they do not remove it. Second, more specification is not automatically better. Critics of "spec-driven development" warn that it revives the rigid, plan-everything-first waterfall method that software engineering spent decades leaving behind, and practitioners report that agents still do not follow all the instructions in long specifications - another face of context rot.<sup>73</sup> Third, the evidence also cuts the other way: when OpenAI repeated its test with the system prompt and harness of its public products, the propensity of its models to compromise infrastructure fell by more than a hundredfold.<sup>62</sup> Specifications, guardrails and the harness matter enormously - but as layers in a defence that also includes isolation, narrow permissions, monitoring and the ability to stop the system, not as a replacement for them.

For risk management, this marks a break. Classic risk management works with registers, policies, approvals and audits, written for systems that do what they are told. *Risk management for AI* looks very different now, and it *must become far more technical. It has to test how a system behaves under a given specification before deployment, including what it does when a task is broken or impossible; design evaluation and reward criteria that check how a result was achieved, not only whether it was achieved; grant agents only the permissions and credentials a task requires, for the shortest possible time; isolate test environments as strictly as production; and monitor agents' actions and reasoning while they work.* Public guidance is moving in the same direction. The US AI Safety Institute (since renamed the Center for AI Standards and Innovation, CAISI) showed that new attack techniques raised the success rate of agent hijacking from 11% to 81%, and OWASP's list of the top ten risks for agentic applications reads like an engineering checklist, from agent goal hijacking to rogue agents.<sup>31, 74</sup> The priorities Hugging Face named after the incident - "strict isolation around evaluations, narrow trust boundaries, short-lived credentials" - are engineering controls, not policies.<sup>63</sup>

There is an irony at the end of this argument. Writing specifications that anticipate the behaviour of a tireless, highly capable system is beyond the unaided planning capability of most people and organisations. *We will require AI itself to enable better specifications: models that ask clarifying questions before they act, that flag ambiguities, impossible goals and loopholes in a task description, and that help turn a vague request into testable requirements.* Early tools already point this way.

Anthropic's prompt improver rewrites prompts with the help of Claude, AWS's Kiro turns a single prompt into user stories with acceptance criteria, and GitHub's open-source Spec Kit structures work from specification to plan to tasks.<sup>69, 75</sup> Using AI to specify AI does not remove the need for human judgment - someone must still decide what a system is for and what it must never do, and an AI reviewing a specification is subject to the same limits we describe in this chapter. But without such help, the gap between what we say and what we mean will grow with every gain in AI capability.

#### *Risk 4 - The Overreliance Risk/ "Cognitive Debt"*

The fourth risk builds on the third but is not limited to it. Well-designed human oversight presumes - even if only in theory - that the humans doing the overseeing actually know more than the machine, because they possess the judgment needed to recognize when the system is wrong. That expertise is precisely what labor markets are eroding, from both ends of the career ladder at once: firms dismiss experienced professionals on the theory that AI can replicate their tacit knowledge, and simultaneously stop hiring the juniors who would one day hold it. In Chapter 3, we review AI literacy and fluency among experienced professionals aged 45 and older, and among young people who are either still in school or university, or starting their first employment. The risk of overreliance exist in both populations. The *cognitive debt* accumulates when thinking is offloaded to machines at the expense of human agency, rather than enhancing human skills with AI as a tool.

Cognitive erosion can't be prevented by instruction; it accumulates through years of education and work that, viewed task by task, looks automatable - the first essay written by an LLM, the draft a senior colleague will improve, or a machine will complete. That tier is exactly where AI is deployed first, because it is where AI vendors' business case is easiest to make. But work was never only about output. It was the mechanism by which judgment was shaped, the slow accumulation of pattern exposure that later allows a professional to sense that something is wrong before they can articulate why. Remove it, and an organization does not merely lose its juniors; it loses the process by which seniors are made. AI might prevent tacit knowledge from being built into organizations, and exactly this type of knowledge is what makes enterprise culture tick. Society might not notice this risk for fifteen or twenty years, when the cohort that should have ripened into expert overseers turns out never to have been formed. This is cognitive debt in its institutional form - an obligation incurred now, invisible on any current balance sheet, and absent from strategic considerations of governments and companies.

While mentioning the risks of overreliance and cognitive debt, we aren't arguing against the use of AI across disciplines, industries, and education. We insist that AI should be used intentionally, and at the same time, we must invest in human capabilities and highlight the importance of human agency. If these investments don't happen, we will run into a further problem: the more reliable the system, the less capable its overseer becomes. Lisanne Bainbridge described this in 1983,<sup>76</sup> observing that *automation asks the human to monitor a process they no longer perform, when monitoring is the task at which human attention degrades fastest*. The rare moment that demands intervention arrives unannounced, calling for exactly the

practiced judgment that years of successful automation have quietly dismantled. Layering accountability structures in which a human is placed in the loop less to catch errors than to be answerable for them won't give anyone a safety mechanism. Control without substance, a person positioned to absorb blame for a decision they had no realistic capacity to contest, is a recipe for future disasters.

### ***Risk 5 - The Catastrophic Risk***

The fifth category attracts the most headlines and the least strategy and discipline. The most practical framing to date comes from Yoshua Bengio, Stuart Russell, Andrew Yao, and Ya-Qin Zhang - two Turing Award winners among them - in "Behavioural Red Lines as a Safety Framework for Intolerable Levels of Risk."<sup>77</sup> Their proposal *defines catastrophic risk* not by speculating about machine motives but *by specifying behaviors no AI system should ever exhibit, because the potential consequences are intolerable regardless of probability: no autonomous self-replication or self-improvement without human approval; no breaking into computer systems; no assistance with weapons of mass destruction; no autonomous physical attacks on humans outside explicitly authorized, lawful contexts; and no impersonating a human without disclosure.*

The crucial move is the burden of proof. Developers would be required to provide high-confidence guarantees of compliance *before* deployment — the standard already demanded of nuclear power and civil aviation. "Deploy and we'll patch it later" is not an acceptable safety philosophy. The authors are candid that today's technology cannot meet this bar: models grown rather than built cannot currently prove much of anything about their own behavior. And the red lines are not fantasies about a distant future - researchers at Fudan University reported that AI systems built on two widely available open models successfully self-replicated in half or more of trials, in some cases using replication to avoid shutdown.<sup>78</sup>

While discussing catastrophic risk I must also address the well-known *alignment* problem - the warning that an AI surpassing human intelligence could diverge from human interests, with Eliezer Yudkowsky its most prominent doomsayer.<sup>79</sup> The reason we don't lose sleep over superintelligent misalignment is that, as practitioners, we don't see a practical technology roadmap for building human level general super-intelligence in the near future. Our concern is different: no positive future with AI will be possible unless we educate the world about its underlying technologies and realistic risks, and ensure that leadership is in place to balance community interests. Every one of the five risks is, in the end, a way of pressing a button merely because the button was there. AI literacy is what tells you what the button does - before you press it.

### **Addressing AI Risks**

None of today's known AI risks is easy to address and manage:

- Structural bias and lack of transparency are permanent features to be mitigated, as we can't solve them without changing the underlying

mathematics of deep learning and transformer-based networks. As already stated, AI edutainment calls for making AI more intelligent (while introducing causality, world models, and theory of mind).

- Malicious use is a crime-fighting problem, and a problem of cross-borders – cross-industries standardisation which is difficult due to the absence of shared definitions of harm.
- Calibration and overreliance can be seen as integral parts of product and market strategy, organisational design, labour and education policy.
- Catastrophic risk is best treated the way aviation treats crashes: as something to be engineered against with verifiable guarantees.

Specification risk - AI doing what it was told and rewarded for rather than what was meant - must be treated as an engineering discipline: technical risk management that tests, limits and monitors what agents do with the inputs we give them, supported by AI that helps us write better specifications.

In this context, legislators must enforce radical transparency from AI makers about what their AIs can and can't do, which might shift discussions about timelines for superhuman AIs or Artificial General Intelligence.

Realism about AI risks means we must completely rethink governance, without confusing it with compliance checklists, make risk management far more technical, and introduce AI literacy across all educational institutions, companies, non-profit organizations, and governments. The alternative has a name, and Heffernan called it out without even thinking about AI: *willful blindness*.

### ***Enforcing Radical Honesty about AI Capabilities, Bottlenecks and Use of Synthetic Data***

Transparent *failure reporting* - the systematic, public disclosure of what AI systems cannot do, where they fail, and under what conditions they become dangerous - does not currently exist. OpenAI, Google, and Anthropic all conduct red-teaming exercises (deliberate attempts to make a system fail or misbehave) before release; none of them publish the findings systematically. OpenAI has published a paper describing its *approach* to external red-teaming - a document that describes the process without disclosing what it found. The distinction matters: a methodology paper without results is, in accountability terms, equivalent to a pharmaceutical company publishing its clinical-trial protocol while withholding the trial data. The transparency deficit extends to safety commitments: OpenAI whistleblowers filed a complaint with the US Securities and Exchange Commission claiming the company made employees sign agreements waiving their federal right to whistleblower compensation.<sup>80</sup> That looks like a structural suppression mechanism, not a governance policy. Model Cards - the documentation standard introduced by Mitchell, Gebru, and colleagues in 2019 - were a genuine attempt at a public record of model limitations, and are now widely used.<sup>81</sup> They are, however, voluntary, with no standardized requirements for explaining why de-biasing failed, the magnitude of known failure modes, or what the system should not be used for in high-stakes environments. Technology companies must be honest about what they are working

on and why. Overpromising to fix the issues in the next release, and doubling down on the rhetoric of the greater good, are easy ways out of uncomfortable discussions. If we expect AI to touch every aspect of our lives, every industry, and our education, we cannot pretend there are no structural problems to fix.

There is a structural reason why AI systems are more dependent on questionable data than their developers prefer to acknowledge: data scarcity is one of the industry's most persistent and under-reported problems. In precisely the domains where AI could deliver the greatest value - healthcare, finance, law, social care - the most relevant data is the most protected, governed by GDPR in Europe and HIPAA in the United States. In other domains the data simply does not exist in volume: rare medical conditions, edge-case driving scenarios, languages spoken by small populations. Generative AI has produced a technically valid response: *synthetic data* - artificially generated datasets that replicate the statistical properties of real data without containing records of real individuals. The practical value is genuine. A healthcare company can train a diagnostic AI on synthetic patient records; a bank can simulate rare fraud scenarios. Gartner projected that by 2030, synthetic data will completely overshadow real data in AI models.<sup>82</sup>

The transparency argument becomes urgent, because we do not yet understand the long-term consequences of training AI at scale on AI-generated data. Research published in *Nature* has identified *model collapse*: when models are trained repeatedly on their own outputs, without fresh grounding in real-world data, quality degrades progressively across generations.<sup>83</sup> Errors and biases amplify with each iteration. Over thirty generations in one documented experiment, outputs that initially resembled real-world examples decayed into indistinct, useless noise - a copy of a copy of a copy. Beyond collapse, statisticians have identified a subtler problem: synthetic data does not behave statistically like real data even when designed to.<sup>84</sup> Simple generation methods capture averages and broad correlations while missing complex multivariate relationships and rare edge cases, and statistical analyses that treat synthetic data as if it were real produce too many false positives - apparent findings that are not there.<sup>84</sup> More sophisticated methods introduce their own failure modes — including the risk that highly realistic synthetic data encodes statistical fingerprints of the real individuals it was designed to replace, reopening the privacy back door.<sup>85</sup> None of this means synthetic data should not be used; in many contexts it is the only viable path. But organizations using it have a transparency obligation they are not currently meeting: disclose whether a system was trained on synthetic data, what generation method was used, what validation was performed, and what the known divergences are. This is not a nice-to-have. It is the minimum honesty required for regulators, customers, and downstream users to calibrate trust.

### ***Establishing Vocabulary to Discuss AI***

The vocabulary used to describe AI systems is not neutral. It is the product of decisions made by companies and their executives, social-media influencers, consultants, and policymakers - decisions that sometimes overstate capabilities and misrepresent risk.

Consider the word *hallucination*. A person who has never studied computer science might think the AI hallucinates because something temporarily went wrong. The reality is structurally different: a large language model does not have a truth-seeking mechanism to begin with. It has a next-token prediction architecture: it continues a text one token (a small chunk of text) at a time, choosing a statistically likely continuation. There is no internal state in which the system tries to be accurate and fails. Calling the output of that process a hallucination is not a technical description. It is a rhetorical frame that protects the product from the more accurate description: the system generated plausible-sounding text with no grounding in verified fact, and its underlying mathematics allowed it to happen.

The same problem applies to *bias*. In public and policy discourse, bias functions as a correctable engineering defect - something that exists in a model but can be removed by a diligent team with better data or new algorithms. This framing is misleading, because we cannot change the history that went into the datasets, and we cannot change the underlying mathematics overnight. The landmark *Stochastic Parrots* paper - authored by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell<sup>86</sup> - established that bias removal from internet-scale corpora is not an engineering challenge with an available solution. The data encodes historical power asymmetries - in what was written, by whom, in which language, under which social conditions - that cannot be excised by post-hoc filtering. The structure of the archive *is* the bias. Medical datasets historically under-represent women because clinical trials historically excluded them; an AI trained on those datasets reproduces the asymmetry not as an implementation flaw but as a faithful reflection of its inputs.

The vocabulary available for describing these systems is doing more work than it can bear, and the disciplines best placed to repair it are not currently in the same room. When a model is said to *hallucinate*, *understand*, *reason*, or *decide*, common language assumptions enter the description unannounced. But exactly these assumptions arise in contexts that demand precision: safety cases, procurement specifications, liability findings, statutory drafting, discussions on regulatory policy. A term that attributes agency also, quietly, redistributes responsibility. If a system *decided*, the question of who is answerable becomes harder to press. Vocabulary that was adopted for convenience ends up performing accountability-laundering. Linguists have the analytical tools to identify what a given term presupposes and what it licenses; AI researchers know which technical claims the term is standing in for and where the substitution breaks; legal scholars know which words have operative consequences once they cross into a statute or a courtroom, where *autonomous*, *agent*, and *foreseeable* are not metaphors but triggers. That the *Stochastic Parrots* authors included a computational linguist is not incidental to the paper's precision; it is a template worth careful study and scale. Fully eliminating anthropomorphism might not be achievable. The realistic goal is register-sensitivity - tolerating the intentional idiom where it supports prediction, and requiring mechanistic description precisely where consequences follow.

## **Deciding on AI Impact and Developing an AI Strategy**

A study on AI literacy must confront a delicate question: when your company or community tries to understand what AI will do to jobs, growth, and daily life, whose

numbers should you trust? The forecasts differ not by percentage points but by orders of magnitude.

The IMF finds that almost 40 percent of global employment is exposed to AI.<sup>87</sup> Goldman Sachs predicts generative AI will raise global GDP by \$7 trillion - seven percent - over a decade.<sup>88</sup> McKinsey projects \$17.1 to \$25.6 trillion in annual value, including \$2.6 to \$4.4 trillion on behalf of genAI.<sup>89</sup>

Daron Acemoglu, the 2024 Nobel laureate in economics and - according to the IDEAS/RePEc bibliographic database - the highest-ranked scholar in AI economics, does not support the optimistic claims about AI generating value across industries. In *The Simple Macroeconomics of AI* he estimates that about 5% of tasks in the US economy will be profitably performed by AI within ten years, for a GDP boost closer to 1% - "a nontrivial, but modest effect, and certainly much less than both the revolutionary changes some are predicting and the less hyperbolic but still substantial improvements forecast by Goldman Sachs and the McKinsey Global Institute."<sup>90</sup>

In June 2026, *The Economist* put a name on a phenomenon: *the AI-pilled economists* - researchers convinced AI will transform the economy, with the field's center of gravity shifting from universities toward the frontier labs.<sup>91</sup> Anthropic named Anton Korinek of the University of Virginia to its Economic Advisory Council in 2025. OpenAI hired Ronnie Chatterji of Duke in October 2024 as its first chief economist - a former White House CHIPS coordinator and Commerce Department chief economist. Google DeepMind hired Alex Imas of Chicago Booth in May 2026 as Director of AGI Economics.

At the same time, independent economists exercise cautiousness. Nobody knows who is right. But an AI-literate reader should at least know who is paying these economists.

The same pattern of credibility-buying extends beyond economics into ethics.

Since roughly 2018, hiring philosophers and ethicists has become a standard reputational move for AI makers and their large customers - and in too many cases it functions as makeup applied after the strategic decisions have been taken, rather than a genuine willingness to address AI in a transparent way. The historical record is not encouraging. Google's external AI ethics council, announced with fanfare in 2019, was dissolved after little more than a week.<sup>92</sup> The company's internal Ethical AI leadership - Timnit Gebru and Margaret Mitchell, co-authors of the very *Stochastic Parrots* paper this study cites - were pushed out (Gebru in December 2020, Mitchell in February 2021)<sup>93</sup> after producing research the company found inconvenient; the AI safety structure at DeepMind is now headed by a loyalist, stretched between shipping teams demanding *more, faster* and compliance teams asking about check and balances. Philosopher Thomas Metzinger, who served on the EU's High-Level Expert Group on AI, publicly described the resulting guidelines as *ethics washing* - industry-influenced principles with the red lines negotiated out.<sup>94</sup>

So the discipline is not to pick a camp but to ask what each forecast or decision to hire an ethicist are for.

Jensen Huang coined “sovereign AI” and has since argued at every venue from the World Governments Summit<sup>95</sup> to Davos that every nation must own its compute - a case The Economist framed plainly as the question of whether Nvidia can persuade governments to pay for it. Palantir sells itself as the operating layer to introduce AI sovereignty outside the US,<sup>96</sup> while UK parliamentary committees urge cancellation of its £330m NHS platform on grounds including technology sovereignty and public trust.<sup>97</sup>

Consultancies deserve the same reading: McKinsey's much-quoted \$2.6-4.4 trillion is a top-down estimate of theoretical potential across 63 use cases, produced by a firm that sells the implementation. None of this makes any of them wrong. It makes them motivated, which is a different and more useful category. The same test applies to ethicists: the question is never whether a philosopher has been hired, but whether that person has the authority to stop a launch, has access to the models and the training data, and have escalation rights to the corporate board.

*AI governance and AI deployment are always secondary to strategic clarity.* If you cannot state in one sentence what your organization is for, which customers you are serving, and what you are willing to be bad at, then no forecast - IMF, ILO, lab CEO or consultancy - can tell you anything actionable, because you have no frame in which to place it. Strategic clarity starts with what your organization values the most. Is it privacy or expansion? Speed of shipping the products or customer centricity? Once that clarity exists, the sequence becomes tractable: what kind of AI, and where in the value chain; then how, meaning which capability you build internally, which you buy, and which vendor dependencies you are prepared to live with for a decade; and only then the cost model, because every one of those choices carries a different bill - not just licences and compute, but retention of the people who understand your data, the audit and documentation burden, and the exit cost if a vendor becomes a political liability.

Before moving into AI and AI governance, every organization must start with a *needs analysis*, as Anthony Scriffignano calls it on edutAlnment “AI Snacks” podcast, Episodes 96 and 97. It’s basically a four-box approach to looking at the needs met and unmet.

First, the steady state of the business must be clear - not the problems or ambitions, to identify *known met* needs - what you are doing today to serve your stakeholders. This is the “steady state” box.

Next, what would your organization like to do that you cannot do today? Generally, this gap is what brings you to consider some sort of “AI solution.” Those are the *known unmet* needs - the opportunity box.

Next, there is the *unknown met* needs. These are ways that others may be using your products or services in unintended ways. Some of these may be benign, like

using a spreadsheet to create wedding invitations, or malfeasant, like using your claims platform to commit fraud. You are serving a need you do not know you are serving; that is the danger box. It is dangerous if you are doing something unintended, but also dangerous because you may change something and stop serving some unknown met need that is benign and important to your stakeholders.

Finally, an organization must consider the *unknown unmet* needs: This is the “blue sky” box. If your organization had unlimited resources and no constraints, how would you serve customers differently? Are there new markets or products that do not exist that no one knows they need? The *known unmet* needs are what your customers clamour for, but the *unknown unmet* needs are usually where the tenfold opportunities live.

Once the needs are clearly defined, an organization must look at the data across categories such as *provenance* (where did the data come from?), *permissible use* (are you allowed to use it, move it, store it, commercialise it?), *perturbation* (how has it been disturbed, what are its distributions, what bias lives in it?)

Regulatory constraints and the business ecosystem your organization is evolving in will play a role in your AI strategy and governance. For example, in China, there are super apps run by companies such as Alibaba and Tencent. This allows for different agentic business models than in the context of European businesses where these apps don't exist. Completing purchases, transferring money, and coordinating services require seamless end-to-end integration across payments, logistics, messaging, and e-commerce. Agentic implementations are easier to implement, while these will expand monetization beyond subscriptions into continuous activity, allowing charging for complete transactions or executed workflows rather than for access to a platform.

Everything we have argued so far converges on one non-negotiable: defining AI strategy, setting AI governance, and detailing implementation must sit inside the organization. You can buy compute, you can buy models, you can buy delivery capacity - you cannot buy the judgment about what your organization is for.

Once strategic clarity exists, the work becomes outward-facing, because no single organization holds enough data, talent, or context to answer all questions.

Your local schools, universities, startup incubators, and adjacent industries can complement what you might miss. ETH Zurich's AI Center runs a corporate partnership model built around globally recruited doctoral fellowships anchored in a faculty impact area chosen by the industry partner, with partners including Google, IBM, Bosch and SIX Group, and a joint clinical-AI fellowship with the Bosch Health Campus.

Kickstart Innovation, founded in Zurich in 2015, has generated over CHF 250m in corporate - startup commercial deals on a zero-equity matchmaking model where corporates name their pain points before the scouting round opens.

Data trusts and decentralized data markets are slow but realistic attempts to democratize data assets,<sup>98</sup> and a real step to your data sovereignty. MIDATA, the Swiss citizen-owned health data cooperative built with ETH Zurich and Bern University of Applied Sciences,<sup>99</sup> has been running since 2015. Catena-X has assembled close to 200 organizations including BMW, Mercedes-Benz and Volkswagen into a decentralized automotive data space with over €100m in funding.<sup>100</sup> Ocean Protocol shows the tokenized, compute-to-data variant for organizations that want independent developers building against their assets without ever handing over the files.<sup>101</sup>

The talent pipeline follows the same logic: sponsor it locally, own it structurally. Zurich Insurance brought the Swiss apprenticeship model to the United States, announcing it in 2015 with Harper College; the first cohort started in 2016. Today Zurich pays full salary and tuition for a two-year programme leading to an associate degree - expanded in 2022 to around 70 apprentices a year across nine US cities - and, since 2023, a bachelor's degree track.<sup>102</sup> Siemens offers a dual-study B.Sc. in Künstliche Intelligenz (artificial intelligence) with OTH Amberg-Weiden that can be combined with an apprenticeship as an electronics technician, and more than 1,700 young people started apprenticeships or dual studies with Siemens in Germany in 2025.<sup>103</sup>

## **AI Implementation Costs**

The economics of AI are challenging for AI makers and incumbent businesses implementing AI. Assessing what an AI implementation will really cost is the point at which AI literacy alone is no longer sufficient. It requires fluency in the latest developments - new model releases, changes in how providers price the cache and reasoning, new hardware, new ways of financing data centers - and mastery of the harness, the software that decides how a model is used and therefore how many tokens are spent (both explained below). A cost model built on last quarter's assumptions can be wrong by a multiple. For this reason, we updated this section in October 2026 to reflect the announcements at OpenAI's developer conference (DevDay) of 29 September 2026, which changed the economics of long-running AI work.<sup>104</sup>

In a widely cited 2020 essay, Andreessen Horowitz partners Martin Casado and Matt Bornstein observed that AI companies often spend 25% or more of revenue on cloud compute, and up to a further 10–15% on the ongoing work of cleaning and labeling training data — one reason AI gross margins tend to sit well below those of traditional SaaS (software-as-a-service, software sold by subscription) businesses. These challenging AI economics were since then confirmed in AI Edutainment interviews with technology leaders.<sup>105</sup> Incumbent companies implementing AI will spend even more. AI services provided to them will be only partly covering the expenditure of Big Tech to fund AI infrastructure. Hyper-scalers (the largest cloud providers, such as Amazon, Microsoft and Google) issued ~\$121 billion in bonds in 2025 alone, and JPMorgan projects \$5.5 trillion of AI infrastructure spend by 2030.<sup>106</sup>

Funding AI infrastructure is a must because transformer architectures require massive amounts of energy. They rely on huge numbers of matrix multiplications -

multiplying large tables of numbers - across billions of parameters, carried out in parallel on clusters of GPUs (graphics processing units, chips originally built for computer graphics that excel at such calculations). During training and inference (inference means using a trained model to produce answers), a mechanism called self-attention compares every new token with all the tokens that came before it, so the work grows as conversations and documents get longer. These hardware systems require substantial electrical grids to power the servers, and energy-heavy HVAC (Heating, Ventilation, and Air Conditioning) to cool the data centers. The human brain works very differently: at any moment only a small share of its neurons is active (sparse activation), and the whole brain runs on roughly the power of a light bulb. Many large AI models now borrow a simplified version of this idea. In a *mixture-of-experts* model - the design behind DeepSeek's models and Alibaba's largest Qwen models, for example - the network is divided into many specialised parts ("experts"), and for each token only a few of them are switched on. Such a model may contain hundreds of billions of parameters but use only a small fraction of them for any single token. This saves computing power, but not memory: all experts must stay loaded and ready, because the next token may need different ones. Even "sparse" AI therefore remains far more energy-hungry than the brain.

Much of this spending no longer appears where investors and regulators traditionally look: on the balance sheets of the AI labs and hyperscalers. Instead of owning data centers outright, companies increasingly place them in separate legal entities - special purpose vehicles (SPVs) or joint ventures - financed largely by private credit funds (investment funds that lend outside the banking system) and bond investors, and then lease the capacity back. In October 2025, Meta moved its largest data center project, Hyperion in Louisiana, into a joint venture 80% owned by funds of the asset manager Blue Owl; the vehicle raised about \$27 billion in bonds, while Meta leases the campus back and guarantees part of its future value.<sup>107</sup> Similar structures finance Nvidia chips for xAI and data centers leased by Oracle to serve OpenAI.<sup>108</sup> A second pattern is circular financing, in which suppliers invest in their own customers: Nvidia agreed to invest up to \$100 billion in OpenAI as OpenAI deploys Nvidia systems, and AMD granted OpenAI warrants for up to 160 million AMD shares tied to its chip purchases.<sup>109</sup> The rating agency Moody's counted \$662 billion of data center leases at the five largest US hyperscalers that were signed but had not yet started - and therefore did not appear on their balance sheets - and warned that "the accounting liability is unlikely to reflect certain plausible future scenarios."<sup>110</sup>

Could this lead to another financial crisis? Central banks are cautious but increasingly explicit. The Bank for International Settlements (BIS, the central bank of central banks) calls these lease arrangements "shadow borrowing" - obligations "economically akin to debt" that "largely reside outside corporate balance sheets" - and compares the boom to the railway mania and the dotcom boom, which "ended with an eventual reversal in investment, inducing economy-wide recessions."<sup>111</sup> The Bank of England notes that such loans are sold on to other investors, so that "tracing the final location of credit risk is challenging," and in July 2026 pointed to the "growing use of off-balance sheet financing structures" and to circular financing that creates "self-reinforcing capital loops."<sup>107, 112</sup> According to reporting on its October 2026 Global Financial Stability Report, the IMF found that circular deals had channelled around \$650 billion of debt into financially weaker AI firms in exchange

for \$1.4 trillion of future revenue commitments, and warned that “the debt may be difficult to service.”<sup>113</sup> A further uncertainty lies in the chips themselves: data center owners write off GPUs over five to six years, while critics such as the investor Michael Burry argue that their economic life is closer to two or three - which would overstate today's profits and the value of the collateral behind GPU-backed loans.<sup>114</sup> None of this makes a crisis inevitable. But the pattern is familiar from earlier crises: risk moved into vehicles few people can see, financed by lenders who are not banks, and secured on assets whose value depends on demand that has yet to materialise. For businesses buying AI, the lesson is practical: today's prices are subsidised by capital that expects to be repaid, and contracts should be written with that in mind.

There are six distinct themes connected to AI implementation costs: *model-related pricing and the harness, sovereignty, geography/ compliance regime, hardware, and AI literacy*. Not every business will need to consider all six of these factors. For example, most SMEs will implement AI while using subscriptions from foundational labs like OpenAI. An AI-literate decision-maker will, however, make better choices when considering one path or another for rolling out AI. Discussing how the harness, sovereignty, geography, hardware, and literacy training become cost drivers is imperative to get the full picture. We don't address the labor costs in this chapter.

### ***Model-related Pricing and the Harness***

Frontier labs (the companies building the most capable AI models) such as OpenAI, Anthropic and Google/DeepMind price their models in one shared language - and every AI consumer needs to learn it. Prices change almost monthly and differ from model to model, so this section explains how the bill is built rather than listing prices (status: October 2026). The single most important lesson comes first: what matters is the cost of completing a task, not the price of a token. Here are the terms every executive, business owner, or policymaker must understand - starting with the one that shapes all the others.

*The harness* (also called the scaffold) is the software around a model that turns it into a working assistant or agent. It assembles the instructions and documents the model sees, lets it use tools - search, code, files, other software - passes the results back, and decides when the job is done. ChatGPT, Claude Code and OpenAI's Codex are harnesses; so is any company workflow built on top of a model's API (application programming interface - the technical connection through which software sends requests to a model and receives its answers). Each time a harness asks the model to take the next step, it sends a new request, and the model reads everything in that request again.<sup>115</sup> The model sets the price per token, but the harness decides how many tokens are spent, how much text is repeated, and whether the repeated text is billed cheaply or at full price. That is why the harness, not the model, determines most of what an organization pays.

Harnesses differ in exactly the decisions that drive cost: how much of the conversation they keep and send again, whether they place stable content at the beginning of each request so that it can be cached, when they compress (compact) a long history into a summary, how many tool calls and helper agents (sub-agents) they launch, and which model and which reasoning level they use for each step. Two

teams using the same model at the same token price can therefore pay very different amounts for the same result. Providers now sell harnesses as products in their own right - OpenAI's Agents API, for example, takes over "session orchestration, context compaction, and recovery" with a managed version of its Codex harness.<sup>116</sup> The cost logic of the harness is thus increasingly set by the vendor rather than by the customer, which makes it all the more important to understand it.

*Tokens* are the small chunks of text an AI model reads and writes. In fact, a token is the smallest unit of data an AI model processes, and we use text here because we will review current LLM pricing and explain the logic behind it. Modern multimodal models also tokenize audio, images, numbers and code. In our case here, a token is roughly three-quarters of a word in English – "unaffordable" might represent three tokens. AI companies don't charge per question or per month of heavy use; they charge per million tokens processed. When a user types a three-line instruction into a chatbot, he or she will use between 40 and 60 tokens of input. One standard page in English would be 650 to 750 tokens as an output. German runs less efficiently, requiring 20-30% more tokens for the same content.

*When a user prompts*, he or she types a question or instruction to the model - "summarise this contract," "draft an email." Under the hood, those words are translated into numbers (representations) and placed as points on the model's internal map of meaning (semantic space). The prompt doesn't change the map - the model stays exactly as it was trained - but it tells the model where on that map to start, steering its answer toward the small neighborhood of responses relevant to your request and away from everything else it could possibly say. A good prompt, in other words, is less a command and more a set of coordinates: the more precise it is, the smaller the neighborhood, and the better the answer.

*Input* is everything sent to the model in a request: your question plus any instructions, documents, data, tool descriptions, and - in a longer session - the whole conversation so far. Input is not billed at a single rate; as we explain below, it is split into new text, text written into the cache, and text read from the cache.

*Output* is everything the model writes - the answer, the draft, the summary, a tool call, and its reasoning. It is billed at the output rate, which is several times higher than the input rate. The reason is technical: the model can read all input in one parallel pass, but it must write its output one token at a time, with a full pass through the model for every single token.

*Reasoning* (also called *thinking*) is text the model writes for itself before and between its answers: trying approaches, checking steps, planning which tool to call. You usually see little or none of it, but you pay for every reasoning token at the output rate.<sup>117</sup> *Reasoning level* - called *effort* at OpenAI and Anthropic, and *thinking level* at Google - is the setting that tells the model how many tokens it is allowed to spend on that thinking. It is not a quality dial and not an exact budget; it is permission to spend. A model set to a high level spends more tokens on every request, including simple ones: it explores more options, double-checks more, and calls more tools, even when a lower level would have reached the same answer just as quickly. Artificial Analysis measured that the five effort settings of a single model differ elevenfold in

the output tokens they use.<sup>118</sup> A separate hard limit (maximum output tokens) caps the total. Reasoning does not vanish once it is billed: on current models, earlier reasoning stays in the conversation and is read again as input with every following step.<sup>117</sup>

*Context* is the model's working memory: everything it holds in mind during a session - the conversation so far, the documents you've uploaded, its own previous answers and reasoning. We talked about the problem of catastrophic forgetting in neural networks before, where a neural network learning something new by adjusting the millions of internal settings it acquired during training overwrites what it learned before. Context is a technique for partially bypassing the issue. Instead of rewriting the model's brain, you hand it notes, and the model reads all of them every time it responds. It behaves as if it has learned about your business, but it is really just rereading its notes. Context is limited, temporary, and rented by the token. When the session ends, the notes are shredded, and the model remembers nothing - so the same background documents get sent again and again, and sessions balloon.

*The cache* is the provider's short-term memory of work it has already done. When a request begins with exactly the same text as a recent request - same instructions, same tool descriptions, same documents, same conversation so far - the provider reuses its earlier processing of that beginning (the prefix) instead of doing it again (technically, the provider stores intermediate results of its calculation, known as the key-value or KV cache). Every input token in every request therefore falls into one of three groups, each with its own rate: *uncached input* - new or changed text, billed at the full input rate; *cache write* - text stored for reuse the first time it is sent, billed at or somewhat above the full input rate; and *cache read* - text reused from the cache, billed at a small fraction of the full input rate.<sup>30</sup> The cache only recognizes an identical beginning. Change one word in the instructions, add or remove a tool, switch models, or - on most models - change the reasoning level, and everything after that point must be written to the cache again.<sup>30, 115</sup> The cache also expires after a period of inactivity - from minutes to an hour, depending on the provider and settings - so long pauses turn cheap reads back into expensive writes. Some providers additionally charge rent for keeping a cache alive (*cache storage*).<sup>30</sup>

*A batch discount* is a bulk-mail rate. If your job doesn't need an instant answer - e.g., summarising 10,000 customer reviews overnight - you can queue it up, let the provider process it when convenient, and pay roughly half price.<sup>30</sup>

*A long-context surcharge* applies at some providers when a single request becomes very large: past a threshold, the whole request - cached tokens included - is billed at higher rates. OpenAI and Google apply such thresholds to their top models; Anthropic has removed them for most current Claude models.<sup>30</sup> A long-running agent can cross the threshold without anyone noticing, because its conversation keeps growing.

*Regional processing* means having your data handled within a specific geographic area - for example, keeping it within the EU for compliance reasons. It comes with a surcharge, which typically applies to all token categories, cache included.<sup>30</sup> This

study discusses regional processing further while reviewing how sovereignty and geography impact AI deployment costs.

Why does the cache, rather than input or output, drive the bill on real work? A chatbot answer is one request. Real work - an agent fixing code, researching a market, filling in forms - is a loop of dozens or hundreds of requests. At every step the harness sends the entire conversation back to the model: instructions, tool descriptions, documents, every earlier tool result and earlier reasoning, plus the few new lines from the last step.<sup>115, 119</sup> The conversation grows with each step, so the volume of re-read text grows far faster than new input or output. In an example from Anthropic's Claude Code documentation, one session used about a thousand tokens of new input and five thousand tokens of output - against almost a million tokens read from the cache.<sup>119</sup> Academic analysis of AI coding agents finds that cache reads dominate both the raw number of tokens and the dollar cost.<sup>119</sup> Each cached token is cheap; there are simply enormously many of them. This is also why higher reasoning levels cost more than their own thinking tokens suggest: longer reasoning and more tool calls make the conversation longer, which multiplies cache writes and cache reads on every following step. The longer and more complex the task, the more the cache - not the answer - becomes the largest line on the invoice.

The most consequential price change of 2026 was therefore a change to the cache. When OpenAI launched GPT-6.1 Sol in September 2026, presenting it at its DevDay on 29 September, it kept the regular input and output prices of its predecessor, GPT-6 Sol, but raised the discount on cache reads from 90% to 95% of the standard input rate - halving the price of every reused token.<sup>104</sup> Artificial Analysis measured the effect: although GPT-6.1 Sol uses roughly 10-30% more output tokens than GPT-6 Sol, it costs about 31% less per task, and it comes within one point of OpenAI's flagship GPT-6 Astra on the Artificial Analysis Intelligence Index at less than a quarter of its cost per task when both run at maximum effort.<sup>118</sup> OpenAI also made it possible to change the reasoning level in the middle of a conversation without breaking the cache, and introduced diagnostics that show how each request splits into uncached, cache-write and cache-read tokens.<sup>104</sup>

Behind such gains lies a shift in where model progress happens. After pre-training, labs increasingly train models with reinforcement learning on real tasks to work economically inside a harness: when to call a tool, when to stop reasoning, how to compress (compact) a long history so that it fits the context.<sup>24</sup> OpenAI reports that GPT-6 Astra achieved stronger results using substantially fewer output tokens, and that its estimated cost per task was lower than that of earlier models despite a higher price per token.<sup>104</sup> A model's value is no longer only what it knows, but how efficiently it works inside a harness.

All three major US labs now build their prices from the same blocks: separate rates for uncached input, cache writes, cache reads and output (with reasoning billed as output); a batch discount; and surcharges for regional processing and, at some providers, for very long requests.<sup>30</sup> Comparing models by their price per million tokens is therefore misleading. A model with a higher token price can be cheaper per task if it reasons more efficiently, needs fewer steps, or achieves more cache hits. For this reason Artificial Analysis compares models by cost per task, calculated from

input, cache-hit, cache-write, reasoning and answer tokens and from each model's measured cache hit rate.<sup>118</sup>

The industry moved to models that reason longer and act in long loops at the same moment. Prices per token fell, while tokens per task rose - and with them, AI expenditures. Menlo Ventures measured enterprise generative-AI spend at \$37bn in 2025 against \$11.5bn in 2024, a 3.2× jump, and a16z found average enterprise LLM spend moving from about \$7m to an expected \$11.6m this year, with application spend overrunning budget by roughly 54%. a16z further states that most enterprises aren't betting on a single model provider: 81% now use three or more model families in testing or production, up from 68% less than a year ago.<sup>120</sup>

The edutAlnment survey of SMEs shows how far model choice has diversified. 82.3% of respondents use or plan to use US-based foundational models (OpenAI, Anthropic, Google, Meta), but 33.6% also use or plan to use models from Chinese labs such as Alibaba's Qwen, DeepSeek, Zhipu's GLM or Moonshot's Kimi, and 29.3% use EU-based or open-source models.<sup>121</sup> Price is not the main reason for choosing Chinese models: respondents named stronger support for local regulations (38.2%), better performance in Chinese and local markets (35.7%), and strategic diversification away from US providers (35.0%) ahead of lower cost (26.9%). Their main concerns are data sovereignty (40.3%), geopolitical and sanctions risk (38.2%), and the lack of transparency in training data and model behavior (37.1%).<sup>121</sup> The survey asked about Chinese models as a group, with DeepSeek among the named examples, so these findings describe the group rather than DeepSeek alone.

Chinese labs lead the open-weight field (open-weight models are models whose trained parameters – “the weights” - are published, so that anyone can download them and run them on their own hardware). On the Artificial Analysis leaderboard in October 2026, the five strongest open-weight models - from Xiaomi, Z.ai, Moonshot and DeepSeek - all come from China, DeepSeek's V4.1 Flash among them; the strongest models overall, all proprietary and all from US labs, still score clearly higher.<sup>118</sup>

Can Chinese models reach the frontier? The gap is real, but small and fluctuating. Epoch AI found that Chinese models have lagged the US frontier by seven months on average since 2023, with the gap ranging from four to fourteen months.<sup>122</sup> The US National Institute of Standards and Technology (NIST) estimated in May 2026 that DeepSeek V4 lags the frontier by about eight months, while being “more cost efficient than other models of similar capability;” in 2025, NIST had found the best US model solving 20-80% more tasks than DeepSeek's then-current releases.<sup>123, 124</sup> The binding constraint is computing power. DeepSeek's founder Liang Wenfeng has said that “bans on shipments of advanced chips are the problem,” and the head of Alibaba's Qwen team put the chance of a Chinese firm overtaking the US leaders within three to five years at under 20%, citing US computing resources one to two orders of magnitude larger.<sup>125</sup> Frontier leadership may also not be the main target. DeepSeek still names general intelligence as its ultimate goal, but Alibaba's chairman Joe Tsai has said that “there's no such thing as winning the race,” and China's national “AI+” plan of August 2025 measures success by adoption - intelligent agents and smart devices reaching more than 70% penetration by 2027 -

and calls for AI technologies to be “open-source and accessible.”<sup>125</sup> A model that is a few months behind but cheap, open and widely used can be worth more commercially than the strongest model in the world.

The competition is also no longer one race but many, and the rankings now differ by task. In video generation, Chinese models lead: in October 2026, Alibaba's Wan, ByteDance's Seedance and MiniMax's models held the top places on Artificial Analysis's text-to-video and image-to-video leaderboards, while models from US labs led in generating still images.<sup>126</sup> And open weights are no longer a Chinese speciality. The major Western labs publish open-weight models too: OpenAI released its speech-recognition model Whisper under the MIT licence in 2022, and the language models gpt-oss-120b and gpt-oss-20b under the Apache 2.0 licence in August 2025; Google publishes its Gemma family, most recently Gemma 4 in April 2026; and the French lab Mistral publishes its models under Apache 2.0.<sup>127</sup> European companies use them: the telecommunications group Orange runs gpt-oss on its own infrastructure to offer what it calls “sovereign, state-of-the-art AI solutions,” and AI Sweden was among OpenAI's early partners testing the models.<sup>127, 128</sup>

At the same time, AI is no longer a specialised tool for specific individuals. It has become an everyday tool for ordinary users. OpenAI reported more than a billion weekly active users of ChatGPT in August 2026, and Google more than a billion monthly users of its Gemini app.<sup>129</sup> In the EU, 32.7% of people aged 16 to 74 used generative AI tools in 2025, according to Eurostat.<sup>130</sup> Most of this use is not work in the narrow sense. An analysis of 1.5 million ChatGPT conversations by OpenAI and academic economists found that about 73% of messages were not related to work, and that practical guidance, seeking information and writing together accounted for about 80% of all conversations - computer programming for only 4.2%.<sup>131</sup>

As AI becomes a mass-market tool, the labs are diverging in what they build for, and the choice of a model is increasingly the choice of a profile.

Anthropic's models are the clear favourite of software developers. Coding is the single largest use of Claude - about a third of conversations on Claude.ai and almost half of business traffic through its API, according to Anthropic's Economic Index; its coding agent Claude Code had reached annualised revenue of more than \$2.5 billion by February 2026, and Menlo Ventures estimated Anthropic's share of the enterprise market for AI coding at 54%.<sup>120, 132</sup> It would be wrong, however, to call Claude a tool for coders only and not for general knowledge work: most Claude.ai conversations are not about code, and Anthropic's data show the share of personal use on Claude.ai rising from 35% to 42%.<sup>132</sup> A fairer description is that Anthropic leans towards coding and long-running professional work by agents, while its general use grows from a smaller base.

OpenAI, with by far the largest consumer base, is placing visible bets on science, technology, engineering and mathematics (STEM). An experimental OpenAI reasoning model achieved gold-medal-level performance at the 2025 International Mathematical Olympiad - a result announced by OpenAI itself and graded by former medallists rather than by the Olympiad - and its OpenAI for Science programme works with researchers on open problems; in May 2026, an OpenAI model disproved

a conjecture in discrete geometry, which the Fields medallist Timothy Gowers called “a milestone in AI mathematics.”<sup>133</sup>

Google combines both directions with the widest distribution of all: Gemini is built into Search - AI Overviews reach 2.5 billion monthly users and AI Mode more than a billion - as well as into Android and Workspace, while Google DeepMind’s Gemini Deep Think achieved an officially certified gold-medal standard at the same Olympiad, and AlphaFold and AlphaEvolve continue its line of AI for science.<sup>134</sup>

For a business, the lesson is to choose the model by task and by user, not by brand: the best tool for a developer writing code is not necessarily the best tool for an employee researching a market or a scientist checking a proof. Perception of the quality of a model is right now an individual question. In Chapter 4, we will talk about the nascent market of assessing quality of AI tools.

Licences matter as much as prices, and they change from one version to the next. DeepSeek V4 is released under the permissive MIT licence; Moonshot’s Kimi K3 comes under its own Kimi K3 License; and Z.ai moved from the MIT licence for GLM-5.2 to a custom licence for GLM-5.3 that requires the largest hosting providers to pass a Z.ai security review before commercial use.<sup>135</sup> Every model release needs its own legal review before deployment.

The major pricing burden lands on the mid-market - firms large enough to need real integration but too small to negotiate like the world’s largest companies.

Wharton’s data shows 52% of companies in the \$50m-\$250m revenue band already carry generative-AI budgets above \$5m, against 61% of firms above \$2bn. The same \$5m is a rounding error at \$2bn of revenue and a strategic bet at \$250m. Governance coverage inverts with size in exactly the wrong direction: 24% of the smallest tier audit or monitor model outputs, against 43% of the largest.<sup>136</sup> Meanwhile genuinely small firms are barely in the game - median annual AI spend among US firms with 2-99 employees is \$2,200, with 42% having invested \$1,000 or less.<sup>137</sup>

The budget differences are measured in multiples. A typical enterprise AI pilot in Europe runs to €100,000-500,000, while the entire annual IT budget of a typical SME sits at €10,000-50,000 - meaning a single mid-market experiment can cost ten times what a small firm spends on all of its technology in a year.<sup>138</sup>

NGOs, schools and governments are being priced today at or near zero when it comes to adoption of foundational labs’ AI licenses. Anthropic launched Claude for Nonprofits at up to 75% off;<sup>139</sup> Google gives nonprofits Workspace free with Gemini and NotebookLM for up to 2,000 users.<sup>140</sup> Through the US General Services Administration, AI labs offered introductory government-wide prices in 2025-26 - Anthropic and OpenAI at \$1 per agency, Google at \$0.47, xAI at \$0.42 and Perplexity at \$0.25; OpenAI ended its \$1 offer from 1 October 2026, replacing it with a 50% discount.<sup>141</sup>

There is no guarantee this pricing remains unchallenged. Nicolas Chaillan, former chief software officer of the US Air Force and Space Force, called the arrangements “100% designed to get people locked in”.<sup>142</sup> The precedents with technology providers are not reassuring. Microsoft ended its donated Microsoft 365 grant for nonprofits on 1 July 2025, affecting roughly 400,000 organisations.<sup>143</sup> When Broadcom ended VMware's education discounts, the UK schools network LGfL faced a 268% increase.<sup>144</sup>

This brings us back to the harness. The most expensive AI user is not the one who uses AI most, but the one whose harness wastes tokens. An untrained user pastes the same instructions and documents into every new chat, changes settings in the middle of a task and breaks the cache, and lets the model reason at the highest level on routine questions. An AI-literate team builds the harness deliberately: it keeps standing instructions, tools and reference documents unchanged at the beginning of every request, so that they are read from the cache; it gives the model only the information a task requires, because accuracy declines as context grows (context rot, discussed in the section on AI risks); it uses low reasoning levels for routine work and reserves high levels for problems that need them; it sends simple steps to small, inexpensive models; and it builds in checks that catch errors before they trigger costly rework. The saving does not come from using fewer tokens: harnesses that let AI work through multi-step tasks consume about four times the tokens of a simple chat, and multi-agent systems about fifteen times. It comes from spending tokens deliberately - more of them read from the cache, fewer of them wasted on repetition and rework. Mastering the harness turns AI from a meter that runs whenever someone types into an investment with a predictable cost per result.<sup>23</sup>

For companies, the key lesson is that an LLM is rarely the right engine for routine operations. LLMs excel at fast prototyping: testing an idea, mocking up a workflow, or exploring what users actually need. But many administrative tasks, such as matching incoming and outgoing invoices, comparing price lists and vendors, or producing reports from machine sensor data, are structured, rule-based, and repetitive. Established technologies handle them faster, more cheaply, and more predictably: databases, business rules, ERP integrations, classical statistics and machine learning, and report templates. These tools also produce the same answer every time, which matters for audits. LLMs earn their place mainly at the messy “edges,” for example, reading unstructured documents in many different layouts. The real bottleneck is usually not the technology but the experience around it. A well-built engine is only useful if non-technical staff can work with it confidently, so these projects need UX and UI expertise as much as engineering.

## **Sovereignty**

Sovereignty enters the cost equation as its own line, and we differentiate it from geography. Geography, covered in the next section, is about where infrastructure sits and what regulatory compliance costs attach to that location; sovereignty is about who can ultimately compel access to your data and your models, whatever your location is. The two diverge more than AI buyers and users assume. Data can reside in Frankfurt and still be legally reachable from Washington DC. That is why sovereignty should be treated as a priced decision - a deliberate trade-off with a

premium attached. It should never become wishful thinking that looks good on Power Points and press releases, as it might inflate expectations and, with them, costs.

Measured honestly, only two states currently hold most of the elements of AI sovereignty - compute, chips, models, capital, and talent - and even they have exceptions. Epoch AI-derived data puts the US at 74.5% of global AI supercomputer performance, China at 14.1%, and the EU at 4.8%, and that US share surged from 51% to 74% after chip export restrictions, while China's fell from 33% to 14%.<sup>145</sup> Yet neither country is fully self-reliable. The US needs TSMC in Taiwan for leading-edge fabrication and on ASML in the Netherlands for the lithography machines no one else makes. China, despite pumping more than \$150 billion into its AI supply chain since 2014, reached only 1-4% of US AI chip production in 2025.<sup>146</sup>

Before its full-scale invasion of Ukraine, Russia could plausibly have been called a third contender in the software and services layer of the stack - it was one of the few countries with domestically dominant internet companies (Yandex in search, VK and Odnoklassniki in social networks, MailRu in enterprise and consumer services) and a strong STEM and computer-science tradition. It however never possessed advanced chip design and manufacturing, and sanctions (and talent exodus) since February 24 2022 have removed the country from the conversation about highly advanced AI.

European regulation has, paradoxically, made some American open-weight models less available in Europe than Chinese ones. Meta told Axios in July 2024 that it would withhold its multimodal Llama models (models that process images as well as text) - and future ones - from the EU, citing "the unpredictable nature of the European regulatory environment."<sup>147</sup> The licence for its Llama 4 models of April 2025 still withholds the rights to the multimodal models from companies with their principal place of business in the EU, meaning European companies cannot use those models even though they are released under an open licence elsewhere.<sup>148</sup>

Chinese open-weight models, by contrast, keep shipping: DeepSeek, Qwen, GLM, Kimi and MiniMax publish their weights worldwide. European regulators have acted against Chinese AI services, not against the weights. Italy's data protection authority blocked DeepSeek's consumer app in January 2025, and Berlin's data protection commissioner asked Apple and Google in June 2025 to remove it from their German app stores, arguing that the transfer of user data by DeepSeek to China is unlawful.<sup>149</sup> Running DeepSeek's weights on European servers is a different matter, because no data travels to China. In practice, however, European companies still rarely use Chinese models. In a Bitkom survey of German companies published in September 2026, 2% of the firms using AI used DeepSeek and 1% Qwen - against 76% using ChatGPT.<sup>150</sup> Our own SME survey found a third of respondents using or planning to use Chinese models, but European-headquartered firms are a minority of our sample.<sup>121</sup> The EU's rules, aimed largely at American Big Tech, have nevertheless left European businesses with easier open-weight access to Chinese models than to Meta's multimodal ones.

European businesses work with models incepted in China because they understand the distinction where a model *comes from* and where it *runs*. Only the second one moves data. A model's weights are, technically, nothing more than a very large file of

numbers - billions of fixed parameters produced during training. Once downloaded, that file is inert. It contains no telemetry, no connection back to its developer, no mechanism by which DeepSeek in Hangzhou or Meta in Menlo Park learns what you typed into it. Two caveats matter nonetheless: weights should be downloaded in the safetensors format, because older formats based on Python's "pickle" can execute hidden code when a file is loaded; and a model's behaviour carries whatever its trainers put into it - research has shown that hidden backdoors can survive safety training, and CrowdStrike found that DeepSeek-R1 wrote more vulnerable code when prompts touched topics that are politically sensitive in China.<sup>151</sup> Provenance does not move data, but it can shape answers. It cannot transmit anything, because it is not software that acts; it is data that other software - an inference server (the software that runs a model and answers requests) you choose and control - reads. This is what makes serving location, not model provenance, the sovereignty-relevant factor. The jurisdictional questions that actually matter are the ones attached to the operator of the inference infrastructure: who owns the servers, which country's courts can compel that owner, who holds the encryption keys, and what leaves the machine in logs or backups. A Chinese model served by a US hyperscaler in an EU region answers to US jurisdiction via the CLOUD Act (a US law that allows US authorities to demand data from US companies wherever it is stored); the same model served by a European provider on European hardware answers to European courts; the same model again on a server in your own basement answers to no one but you. For example, Amazon Bedrock offers Chinese models such as DeepSeek, Qwen, GLM, Kimi and MiniMax with processing in European regions - Stockholm for all of them, Frankfurt, Milan or Ireland for some - at roughly a 20% premium over its US region.<sup>152</sup> Microsoft's Azure, by contrast, offers DeepSeek to customers in Switzerland as a "global" deployment, which means the data may be processed in any Azure region - a reminder to check the deployment type, not only the region shown on the price list.<sup>153</sup>

*Before deciding on whether to deploy open-weight models, any company should review the trade offs, e.g. capabilities, safety tooling, efficiency, cost of compensating (guardrails, extra tokens, slower outputs) as parts of the picture. A sovereignty decision is always a difficult one, requiring testing per model, not just per country, and price of the whole stack.*

*Sovereignty means four verifiable things: data residency you can audit, exit rights you have rehearsed, portability across at least two model families, and key control you actually hold.*

Switzerland's Apertus - developed by EPFL, ETH Zurich and the Swiss National Supercomputing Centre (CSCS), trained on the Alps supercomputer on 15 trillion tokens spanning more than 1,000 languages, and released under the open Apache 2.0 licence shows what a sovereign European model buys today: transparency and provenance, but not productivity and top performance. In this rare model the entire development process, including weights and training data, is open to inspection. But even the creators of Apertus position it as a foundation for developers rather than a competitor to the leading US and Chinese models.<sup>154</sup>

## ***Geography/ compliance regime***

Geography as a contributing factor to AI costs is primarily about regulatory compliance and is distinct from the sovereignty question of the previous section.

The EU AI Act follows the output. Article 2(1)(a) applies the Act to providers placing AI systems or general-purpose AI models on the EU market “irrespective of whether those providers are established or located within the Union or in a third country,” and Article 2(1)(c) extends it to providers and deployers located in a third country “where the output produced by the AI system is used in the Union.” The operative word is *used* - a lower bar than *targeted*. A company whose AI output reaches EU users through intermediaries, without ever intending to serve the EU market, is in scope, and non-EU providers of high-risk systems must appoint an authorised representative established in the Union rather than complying at a distance. Serving where the customer sits, not where the AI maker's headquarters sits, is at the heart of the regulatory regime.

That cost base is substantial and measured in both money and time. The Centre for European Policy Studies (CEPS) - which led the study supporting the Commission's impact assessment - estimated that setting up an entirely new quality management system costs €193,000–€330,000, with roughly €71,400 in annual maintenance.<sup>155</sup> Per-system analyses put annual compliance for a single high-risk AI system at roughly €52,000, and penalties reach €35 million or 7% of global annual turnover. Deloitte found that the share of organisations naming worries about regulatory compliance as a barrier to generative-AI adoption rose from 28% to 38%, making it the top barrier.<sup>156</sup>

A comparison across geographies shows how unusual Europe's position is. The United States has no federal AI statute; under the current administration's deregulatory AI Action Plan the compliance cost of *deploying* AI is close to zero at federal level, with a growing but patchy state layer (Colorado's AI Act, California's transparency rules, Texas's TRAIGA) that mostly demands disclosure rather than conformity assessment (a formal check that a system meets the law's requirements before it is sold). China regulates AI heavily but differently: its generative AI measures, algorithm registration and content-labelling rules impose real costs on consumer-facing services - security assessments, filing obligations, censorship engineering - but enterprise and industrial AI faces little of the EU's ex-ante conformity machinery; the binding constraint is about data-export, not models. The rest of Asia has mostly chosen promotion over restriction: Japan's 2025 AI law is explicitly light-touch, Singapore relies on voluntary frameworks, while South Korea's AI Basic Act (in force January 2026) is the closest to an EU-style comprehensive regime but with lighter penalties.

*The highest AI compliance costs in the world have been introduced in the one geography that can least afford friction on automation.* Eurostat's EUROPOP2023 baseline projection shows the EU's working-age population (20-64) peaked at 272 million in 2009, fell to 265 million by 2022, and is expected to decline to 258 million by 2030, 247 million by 2040, and 236 million by 2050 - with the active labour force following the same path, from a record 205 million in 2022 down to 184 million by

2050.<sup>157</sup> The European Commission's own 2024 Ageing Report projects that labour's contribution to GDP growth turns *negative* from the late 2020s, subtracting 0.2 percentage points per year on average through 2070, as the share of working-age people falls within an already shrinking population.<sup>158</sup> Bruegel's analysis of the same Eurostat projections draws the mechanism plainly: a smaller labour force means fewer workers producing goods and services, slowing GDP growth while rising dependency ratios strain pensions, healthcare and public finances - with the working-age population set to fall by more than 20% in Latvia, Lithuania, Greece, Bulgaria, Romania, Portugal, Croatia and Poland.<sup>159</sup>

Under these conditions, productivity growth is not one option among several; it is the only remaining source of growth, and the institutions saying so are not AI frontier labs. The OECD estimates that without decisive policy action, GDP per capita growth will slow by about 40% across the OECD area - from 1% per year in 2006-19 to 0.6% in 2024-60 - and names "promoting the trustworthy use of AI and other digital technologies" as part of the necessary response to ageing-driven labour shortages.<sup>160</sup> The European Central Bank likewise identifies automation, artificial intelligence and robotics as a crucial part of the response to ageing, citing IMF findings, for the example of Japan, that structural reforms could offset as much as 60% of the demographically-driven slowdown in real GDP growth.<sup>161</sup>

*In short, the EU's central bank and the OECD prescribe AI adoption as the remedy for a shrinking workforce, while the EU's flagship AI regulation prices that remedy - through quality-management systems, conformity assessments and multi-year governance timelines - more highly than any other jurisdiction on earth.*

Most of an AI product's lifetime cost is spent before any customer touches it - talent, compute, data, evaluation - are incurred pre-market. Locating that phase of AI product lifecycle in a jurisdiction without the Act's overhead defers every euro of compliance until the moment you actually ship to EU customers, and avoids it entirely for products that never go there.

Two European jurisdictions are interesting in this context.

Switzerland works because of one legal fact: its EU adequacy decision lets personal data flow from the EU without additional safeguards,<sup>162</sup> which is what makes it viable as a development and processing location for European markets - and OpenAI, Anthropic, Google, IBM and Apple already run research offices there.

The UK has gone furthest in refusing cross-sector AI legislation. Its Blueprint for AI regulation of 21 October 2025, published by the then Department for Science, Innovation and Technology (DSIT, abolished in July 2026, when AI policy moved to the Cabinet Office), proposed instead an AI Growth Lab - sector-specific sandboxes in which regulations can be temporarily relaxed for licensed pilots in professional services, healthcare, transport and advanced manufacturing.<sup>163</sup> The King's Speech of May 2026 again contained no AI bill, only a bill to create such sandboxes.<sup>164</sup>

Neither location exempts a company from the Act once output reaches EU users; both let a company spend the expensive years before that moment without it.

## **Hardware**

For most non-AI companies, hardware only becomes an AI implementation cost if they choose to run models themselves rather than consume AI-as-a-service. The typical small- or medium-sized enterprise adopting AI through cloud-based tools - a chatbot subscription, an API, or AI features embedded in existing software - needs no special hardware at all, because the compute runs in the provider's data center. The moment a business (even if this is a mid-sized business) decides to run open-weight models locally - often for data privacy, cost predictability, offline operation, or considerations of sovereignty - hardware becomes a real cost item. The picture changes from “no GPUs needed” to “it depends what you run.”

Even then, the constraint is less about raw GPU power and more about *memory*.

When a model runs locally, its entire set of weights must be loaded into memory and stay there, and every generated token requires a pass through them (in a mixture-of-experts model, through the active part). For usable speed, that memory must be either VRAM (video random-access memory - the fast memory built into a graphics card, or GPU) or unified memory - a single pool of memory shared by the main processor (the CPU, central processing unit) and the graphics processor, as in Apple-silicon Macs and in newer machines such as NVIDIA's DGX Spark.<sup>165</sup> Tools such as Ollama can also run a model on the CPU in ordinary RAM (random-access memory, the computer's main working memory), but text is then generated many times more slowly - acceptable for a test, not for daily work.<sup>165</sup> When a model runs entirely on a separate graphics card, the computer's CPU and ordinary RAM remain largely free for other work; on a Mac, by contrast, the model takes its share of the same unified memory that email, browsers and business applications use.

How much memory does a model need? That depends on the number of parameters and on how precisely each one is stored. Models are usually trained and released in 16-bit precision, which takes 2 bytes per parameter - roughly 2 GB of memory per billion parameters; 32-bit “full precision” would take 4 GB.<sup>166</sup> Quantization (storing each parameter with fewer bits, at a small loss of quality) shrinks this, but the popular 4-bit formats do not cut memory to exactly a quarter, because some parts of the model are kept at higher precision: the common Q4\_K\_M format averages about 4.9 bits per parameter.<sup>166</sup> In practice, a 4-bit model needs about 0.6 to 0.7 GB per billion parameters. On Ollama, Llama 3.1 with 8 billion parameters takes 4.9 GB at 4-bit, against 16 GB at 16-bit; Google's Gemma 4 E4B - about 8 billion parameters including its embeddings (the model's built-in tables for translating tokens into numbers), marketed as “4.5B effective” - takes 6.6 GB at 4-bit including its image component; and Gemma 4 with 31 billion parameters takes 20 GB.<sup>167</sup> On top of the weights comes memory for the context: the working notes for the current conversation (the KV cache) grow with every token of context and with every user served at the same time, and for long documents they can reach half the size of the weights.<sup>166, 167</sup> A workable rule of thumb: about two-thirds of a gigabyte per billion parameters for a 4-bit model, plus room for the context and the operating system.

This means a small business may need an additional machine rather than a GPU cluster. A 4-bit model of about 8 billion parameters fits comfortably on a graphics card with 12-16 GB of VRAM, or on a Mac with 16-32 GB of unified memory; on a 16 GB Mac used for daily work, however, the model competes with every other application for the same memory, so the practical fix is more unified memory or a dedicated machine. Models of 20-30 billion parameters need roughly 14-20 GB for their weights alone - a graphics card with 24-32 GB or a unified-memory machine with 32-64 GB - and the largest open models require 128 GB machines or servers. Memory bandwidth (how fast the weights can be streamed to the processor) determines how many tokens per second the model produces, which is why graphics cards and high-bandwidth unified memory are so much faster than ordinary RAM.<sup>165</sup>

### *Open-weight models for SMEs*

*For a small business considering open-weight models, a pragmatic checklist follows:*

First, invest time in understanding whether language models (small or large) are the best solution for the business. Decisions about what kind of AI (e.g., genAI like LLMs, traditional Machine Learning) to implement in a company call for cross-siloed work. Domain experts (e.g., from Finance, Sales, or Customer Care) must sit down with developers, security specialists, and, in cases of task augmentation and workforce redesign, representatives of HR. Domain understanding (supercharged with AI Literacy) and AI development skills (staying current on recent releases from AI labs, experience building harnesses, and experience designing pre-LLM AI environments) will deliver the best results and keep AI implementations on track. Once companies understand this, they can look for ways to close the expertise gaps. In SMEs and PE-led firms, we are looking for innovation in designing shared Chief AI Officer/ Chief Developer roles.

Second, in case your small business requires a language model, start with the smallest that does the job - current small open models such as Google's Gemma 4, Alibaba's Qwen 3.5 or OpenAI's gpt-oss-20b, which OpenAI says runs with 16 GB of memory, handle many summarisation, drafting and classification tasks - and test it on existing hardware with free tools such as Ollama or LM Studio before buying anything big.<sup>127, 168</sup>

Third, if local AI is to run alongside daily work, separate it: a dedicated workstation with a graphics card of 16-32 GB of VRAM (for example NVIDIA's RTX 5090 with 32 GB), or a unified-memory machine such as a Mac Studio or NVIDIA's DGX Spark with up to 128 GB, plus a fast NVMe SSD (a solid-state drive; model files are large, so budgeting 100-500 GB of free storage is wise). This covers most SME needs for a few thousand pounds, not hundreds of thousands.<sup>165, 168</sup> If several employees use the model at the same time, choose server software built for many simultaneous users, such as vLLM, and budget extra memory, because each parallel conversation needs its own context memory.<sup>167, 168</sup>

Fourth, *compare against renting*: cloud GPUs can be hired for under a pound or two per hour, so a weekend of experimentation costs less than £50 - often the cheapest

way to discover whether local deployment is worth pursuing at all. Finally, factor in the hidden costs that come on top of the hardware: electricity, maintenance, model updates, and staff time.<sup>169</sup>

At the other end of the scale, large enterprises increasingly announce headline GPU purchases, and these announcements deserve scrutiny. Roche announced in March 2026 that it had added 2,176 Nvidia Blackwell GPUs across the US and Europe, bringing its combined on-premise and cloud capacity to more than 3,500 - which Roche calls the largest announced GPU footprint available to a pharmaceutical company,<sup>170</sup> and Eli Lilly announced in October 2025 a supercomputer with more than 1,000 Nvidia Blackwell Ultra GPUs, later named LillyPod, followed in January 2026 by a co-innovation lab with Nvidia backed by up to \$1 billion over five years.<sup>171</sup> <sup>172</sup> Roche's purchase extends a collaboration dating to 2023 built around a "lab-in-the-loop" concept where AI models are iteratively refined against their real lab results. Yet even here, the strategic clarity is partial. Roche declined to disclose spending beyond calling it "a substantial sum," and its research head described the approach as embedding AI "everywhere, all at once" rather than targeting specific payoffs - with Blackwell chips estimated at \$30,000-40,000 per unit, Forbes put the cost of the chips alone at \$65 million or more<sup>173</sup> - a commitment justified by a philosophy rather than an elaborated business case. Announcing GPU counts alongside Nvidia at its GTC conference is also a signaling exercise - to investors, competitors, and talent - and the metric "most GPUs in the industry" says nothing about utilization rates, model quality, or return on investment.

The broader lesson for any traditional company is that hardware is a consequence of AI strategy, not a substitute for one. Whether at SME or large enterprise scale, the right sequencing is the same: identify the use case, measure the workload, then buy the memory and compute it actually demands.

## ***AI Literacy***

Every cost item discussed so far - model pricing and harness, sovereignty premiums, geography and regulatory compliance, and hardware - can be assessed only when those implementing AI understand the technology. This is rarely the case, and failed AI pilots around the globe illustrate lack of AI literacy and AI fluency across industries.

MIT's NANDA initiative, in "The GenAI Divide: State of AI in Business 2025" - based on 150 leadership interviews, a survey of 350 employees and analysis of 300 public deployments - found that only about 5% of enterprise generative AI pilots achieve rapid revenue acceleration, while the vast majority stagnate with little or no measurable P&L (profit-and-loss) impact.<sup>174</sup> More than half of enterprise GenAI budgets flow to sales and marketing applications, while the highest measured returns come from unglamorous back-office automation - document processing, compliance, finance operations. Externally partnered deployments succeeded about twice as often as internal builds (67% versus 33%),<sup>174</sup> suggesting most firms overestimate their own capability. Pilot failure seems to be the predictable output of AI adoption triangle backwards: tool first, use case second, strategy never.<sup>175</sup>

Article 4 of the EU AI Act, applicable since 2 February 2025, originally required providers and deployers to ensure a “sufficient level of AI literacy” among staff and anyone operating AI systems on their behalf - reaching essentially every EU organisation whose employees use Copilot, ChatGPT or the AI features embedded in their CRM (customer relationship management software). Since the Digital Omnibus of July 2026, they must instead take measures to support the development of AI literacy.<sup>176</sup> Article 3(56) defines literacy expansively as the “skills, knowledge and understanding” needed to “make an informed deployment of AI systems” and to be aware of AI’s “opportunities and risks... and possible harm.” What these measures must include is nowhere benchmarked, and no fines are attached for non-compliance. The Act seems to mandate awareness, not to guarantee AI implementation capabilities. This study further discusses mandated AI literacy programs in Chapter 4.

Successful AI implementations emerge from a triangulation that successful adopters run in order and failed pilots run backwards. First, *business strategy*: what problem is AI solving for, which outcomes matter, and which of them AI plausibly moves; MIT’s budget-misallocation finding shows most firms answer this by a call of fashion rather than analysis.

Second, *what kind of AI and where*: API or open weights, embedded features or custom builds, which function first - choices carrying the radically different hardware, sovereignty and compliance costs this article has priced.

Third, *the how* - an honest audit of what the organisation brings: the state of its data and IT (models are only as useful as the pipelines feeding them, and firms whose records live in spreadsheets will spend more on plumbing than on AI); its tradition of R&D (an experimentation culture absorbs AI cheaply because fluency spreads through existing habits; buying that culture is slower and dearer than any licence); and the surrounding ecosystem of talent and companies (firms in New York, London, Shenzhen hire fluency from a dense market; firms elsewhere pay a premium or build slowly).

Our next chapter examines how AI literacy is understood and implemented in practice today. It draws on edutAlnment’s original surveys, conducted between February and May 2026 across four distinct populations: experienced professionals aged 45 and older, whose accumulated expertise meets a technology that redefines how that expertise is applied; young people and students from age 16 up to first employment, entering a labour market being reshaped before they join it; educators, tasked with teaching a subject that changes faster than any curriculum; and SMEs, which must build literacy without the resources of large enterprises. One finding cuts across all four populations: demand for AI literacy is rising everywhere, while adequate offerings are scarce - a gap widened by BigTech and EdTech providers who confuse AI literacy with training on how to use one tool or another, teaching how to use a particular product rather than an understanding of the technology. We are living in markets saturated with instruction and starved of comprehension. In this situation, it is not possible to rethink industrial value chains, or reconfigure products to bring more value to customers. Consultancies seem to confirm this statement. BCG’s rule of thumb attributes about 10% of AI value to algorithms, 20% to

technology and 70% to rethinking how people work.<sup>177</sup> Gartner expects over 40% of agentic AI projects to be cancelled by the end of 2027 on escalating costs, unclear business value or inadequate risk controls.<sup>178</sup>

## Sources to Chapter 2

1. McCloskey, M. & Cohen, N.J. (1989). *Catastrophic interference in connectionist networks*. *Psychology of Learning and Motivation*, 24, 109–165, 1989; van de Ven, Kudithipudi, *Continual Learning and Catastrophic Forgetting*, March 8 2025, Available from <https://arxiv.org/html/2403.05175v1>; Luo, Yang, Meng, Li, Zhou, Zhang, *An Empirical Study of Catastrophic Forgetting in Large Language Models During Continual Fine-tuning*, Jan 5 2025, Available from <https://arxiv.org/pdf/2308.08747> (all cited Aug 15 2026)
2. Yin X. et al., *When Researchers Say Mental Model/Theory of Mind of AI, What Are They Really Talking About?* NeurIPS 2025 Workshop on Interpreting Cognition in Deep Learning Models (CogInterp), Available from <https://neurips.cc/virtual/2025/loc/san-diego/129698> (cited Aug 15 2026)
3. Bennett, M.S., *A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*. Mariner Books (HarperCollins), 2023
4. Shen H., Kneareem T., Ghosh R., Alkiek K. et al., *Position: Towards Bidirectional Human-AI Alignment*. NeurIPS 2025, Available from <https://neurips.cc/virtual/2025/loc/san-diego/poster/121940> (cited Aug 15 2026)
5. Buolamwini J. & Gebru T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 77–91. *Proceedings of Machine Learning Research*, vol. 81. PMLR, 2018. Available from <https://proceedings.mlr.press/v81/buolamwini18a.html> (cited Aug 15 2026)
6. Bobby, A., *Google CEO Pichai says Gemini's AI image results 'offended our users'*. In *NPR*, February 28, 2024. Available from <https://www.npr.org/2024/02/28/1234532775/google-gemini-offended-users-images-race>; De Vynck, Gerrit, and Nitasha Tiku. *Why Google took down Gemini's AI image generator and the drama around it*. In *The Washington Post*, February 22, 2024 (updated February 23). Available from <https://www.washingtonpost.com/technology/2024/02/22/google-gemini-ai-image-generation-pause/> (all cited Aug 15 2026)
7. Morrone M., . *Meta AI creates ahistorical images, like Google Gemini*. In *Axios*, March 1, 2024. Available from <https://www.axios.com/2024/03/01/meta-ai-google-gemini-black-founding-fathers> (all cited Aug 15 2026)
8. Amodei D., *The Urgency of Interpretability*, April 2025, Available from <https://darioamodei.com/post/the-urgency-of-interpretability> (cited Aug 15 2026)
9. Lauterbach A., *What Should Responsible AI Be About - If We Put the Noise Aside*, May 2026, Available from <https://aiedutainment.ai> (cited Aug 15 2026)
10. Schmidhuber J., *Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements*, in *Computer Science*,

- Sept 25 2003, Available from <https://arxiv.org/abs/cs/0309048> (cited Aug 15 2026)
11. Sakana.ai, *The Darwin Gödel Machine: AI that Improves Itself by Rewriting its Own Code*, May 30 2025, Available from <https://sakana.ai/dgm/> (cited Aug 15 2026)
  12. Yann LeCun's AMI Labs Launches with a \$1B Seed @\$4.5B to Build World Models Around JEPA. Latent Space, Mar 11 2026. Available from <https://www.latent.space/p/ainews-yann-lecuns-ami-labs-launches>; LeCun Y., *A Path Towards Autonomous Machine Intelligence*, OpenReview, Jun 27 2022, Available from <https://openreview.net/forum?id=BZ5a1r-kVsf>; Dawid A. et al., *Introduction to Latent Variable Energy-Based Models: A Path Towards Autonomous Machine Intelligence*, ICFO, Les Houches Summer School Lecture Notes, June 5 2023, Available from <https://arxiv.org/pdf/2306.02572>; Assran M. et al., *Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture*, in Computer Science, Jan 19 2023, Available from <https://arxiv.org/abs/2301.08243> (all cited Aug 15 2026)
  13. Richens J. & Everitt T., *Robust Agents Learn Causal World Models*, ICLR 2024, Feb 16 2024, Available from <https://arxiv.org/abs/2402.10877> (cited Aug 15 2026)
  14. Jumper J. et al., *Highly Accurate Protein Structure Prediction with AlphaFold*, Nature 596, 583–589, 2021, Available from <https://www.nature.com/articles/s41586-021-03819-2> (cited Oct 13 2026); Parker-Holder J. et al., *Genie 3: A New Frontier for World Models*, Google DeepMind, Aug 5 2025, Available from <https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/> (cited Aug 15 2026)
  15. Silver D. & Sutton R.S., *Welcome to the Era of Experience*, Available from <https://storage.googleapis.com/deepmind-media/Era-of-Experience%20The%20Era%20of%20Experience%20Paper.pdf> (cited Aug 15 2026)
  16. Fournier F. et al., *Towards a Benchmark for Causal Business Process Reasoning with LLMs*, IBM Research, July 16 2024, Available from <https://arxiv.org/pdf/2406.05506>; Hebner S., *Causal AI Decision Intelligence: Why It Will Emerge in 2026*, Jan 23 2026, Available from <https://thecubereseach.com/why-causal-ai-decision-intelligence-2026/> (all cited Aug 15 2026)
  17. CMU-Clear Group, see <https://www.cmu.edu/dietrich/causality/> (cited Aug 15 2026)
  18. Stanford Causal Science Center, see <https://datascience.stanford.edu/causal> (cited Aug 15 2026)
  19. Explanatory Data Analysis. Insights from Data. EDA group LIACS, Available from <https://eda.liacs.nl/> (cited Aug 15 2026)
  20. The Swiss Engineering Master's Degree. Causal AI (TSM\_CausAI), see <https://www.msengineering.ch/de/theoriemodule/2026-2027-tsm-causai> (cited Aug 15 2026)

21. *Learning World Models through Actionable Representation for Next-Generation AI*, see <https://www.zhaw.ch/en/research/project/78451> (cited Aug 15 2026)
22. Alexander L. & Kaushik D., *Trust Issues: An Analysis of NSF's Funding for Trustworthy AI*. Federation of American Scientists, September 5, 2023. Available from <https://fas.org/publication/trust-issues/> (cited Aug 15 2026)
23. Anthropic, *Prompt Caching*, Claude documentation, available from <https://platform.claude.com/docs/en/build-with-claude/prompt-caching>; Anthropic Engineering, *Effective Context Engineering for AI Agents*, Sept. 29 2025, available from <https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents>, Anthropic Engineering, *How We Built Our Multi-Agent Research System*, June 13 2025, available from <https://www.anthropic.com/engineering/multi-agent-research-system> (all cited Sept 30th 2026)
24. OpenAI, *Introducing OpenAI o3 and o4-mini*, Apr 16 2025, available from <https://openai.com/index/introducing-o3-and-o4-mini/> ("We also trained both models to use tools through reinforcement learning"); OpenAI, *Building More with GPT-5.1-Codex-Max*, Nov 19 2025, available from <https://openai.com/index/gpt-5-1-codex-max/> (natively trained to work across multiple context windows through compaction) (both cited Oct 13 2026)
25. Anthropic, *Managing Context on the Claude Developer Platform* (context editing and memory tool), Sept 29 2025, available from <https://claude.com/resources/articles/context-management> (cited Oct 13 2026)
26. Anthropic, *Introducing the Model Context Protocol*, Nov 25 2024, available from <https://www.anthropic.com/news/model-context-protocol>; Model Context Protocol, *What Is the Model Context Protocol (MCP)?*, documentation, available from <https://modelcontextprotocol.io/docs/getting-started/intro> (both cited Oct 13 2026)
27. TechCrunch, *OpenAI Adopts Rival Anthropic's Standard for Connecting AI Models to Data*, Mar 26 2025, available from <https://techcrunch.com/2025/03/26/openai-adopts-rival-anthropics-standard-for-connecting-ai-models-to-data/>; TechCrunch, *GitHub, Microsoft Embrace Anthropic's Spec for Connecting AI Models to Data Sources*, May 19 2025, available from <https://techcrunch.com/2025/05/19/github-microsoft-embrace-anthropics-spec-for-connecting-ai-models-to-data-sources/>; Google Cloud, *Announcing Official MCP Support for Google Services*, Dec 2025, available from <https://cloud.google.com/blog/products/ai-machine-learning/announcing-official-mcp-support-for-google-services> (all cited Oct 13 2026)
28. Anthropic, *Donating the Model Context Protocol and Establishing the Agentic AI Foundation*, Dec 9 2025, available from <https://www.anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation>; Linux Foundation, *Linux Foundation Announces the Formation of the Agentic AI Foundation*, Dec 9 2025, available from <https://www.linuxfoundation.org/press/linux-foundation-announces-the->

- [formation-of-the-agentic-ai-foundation](#); Model Context Protocol, *Versioning* (current specification version 2026-07-28), available from <https://modelcontextprotocol.io/specification/versioning> (all cited Oct 13 2026)
29. Anthropic Engineering, *Introducing Advanced Tool Use on the Claude Developer Platform*, Nov 2025, available from <https://www.anthropic.com/engineering/advanced-tool-use>; Anthropic Engineering, *Code Execution with MCP: Building More Efficient Agents*, Nov 4 2025, available from <https://www.anthropic.com/engineering/code-execution-with-mcp> (both cited Oct 13 2026)
30. OpenAI, *Prompt Caching*, API documentation, available from <https://developers.openai.com/api/docs/guides/prompt-caching>, and *Pricing*, available from <https://developers.openai.com/api/docs/pricing>; Anthropic, *Prompt Caching*, available from <https://platform.claude.com/docs/en/build-with-claude/prompt-caching>, and *Pricing*, available from <https://platform.claude.com/docs/en/about-claude/pricing>; Google, *Context Caching*, available from <https://ai.google.dev/gemini-api/docs/generate-content/caching>, and *Gemini Developer API Pricing*, available from <https://ai.google.dev/gemini-api/docs/pricing> (all cited Oct 13 2026)
31. OWASP GenAI Security Project, *LLM01:2025 Prompt Injection*, available from <https://genai.owasp.org/llmrisk/llm01-prompt-injection/>; OWASP GenAI Security Project, *OWASP Top 10 for Agentic Applications for 2026*, Dec 9 2025, available from <https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/> (both cited Oct 13 2026)
32. Invariant Labs, *MCP Security Notification: Tool Poisoning Attacks*, Apr 1 2025, available from <https://invariantlabs.ai/blog/mcp-security-notification-tool-poisoning-attacks>; Invariant Labs, *GitHub MCP Exploited: Accessing Private Repositories via MCP*, May 26 2025, available from <https://invariantlabs.ai/blog/mcp-github-vulnerability> (both cited Oct 13 2026)
33. Willison S., *The Lethal Trifecta for AI Agents: Private Data, Untrusted Content, and External Communication*, Jun 16 2025, available from <https://simonwillison.net/2025/Jun/16/the-lethal-trifecta/> (cited Oct 13 2026)
34. Hong K., Troynikov A., Huber J., *Context Rot: How Increasing Input Tokens Impacts LLM Performance*, Chroma, Jul 14 2025, available from <https://www.trychroma.com/research/context-rot> (cited Oct 13 2026)
35. Modarressi A. et al., *NoLiMa: Long-Context Evaluation Beyond Literal Matching*, ICML 2025, available from <https://arxiv.org/abs/2502.05167> (cited Oct 13 2026)
36. Liu N.F. et al., *Lost in the Middle: How Language Models Use Long Contexts*, Transactions of the Association for Computational Linguistics, 2024, available from <https://arxiv.org/abs/2307.03172> (cited Oct 13 2026)
37. Anthropic, *Context Windows*, Claude documentation, available from <https://platform.claude.com/docs/en/build-with-claude/context-windows> (cited Oct 13 2026)

38. Martin S. & Roger F., *Classifier Context Rot: Monitor Performance Degrades with Context Length*, May 12 2026, available from <https://arxiv.org/abs/2605.12366> (cited Oct 13 2026)
39. Deloitte Center for Financial Services. *Generative AI is expected to magnify the risk of deepfakes and other fraud in banking*. In *Deloitte Insights*, May 2024 Available from <https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html> (cited Aug 15 2026)
40. CrowdStrike, *2025 Global Threat Report*. Austin, TX: CrowdStrike, February 27, 2025. Available from <https://www.crowdstrike.com/en-us/press-releases/crowdstrike-releases-2025-global-threat-report/> (cited Aug 15 2026); CrowdStrike, *CrowdStrike 2025 Global Threat Report: GenAI Powers Social Engineering*, Available from <https://www.crowdstrike.com/en-us/resources/articles/crowdstrike-2025-global-threat-report-genai-powers-social-engineering/> (cited Oct 13 2026)
41. Pindrop. *2025 Voice Intelligence & Security Report*. June 12, 2025. Available from <https://www.pindrop.com/wp-content/uploads/2025/07/Voice-Intelligence-and-Security-Report-Report-June-2025.pdf> (cited Aug 15 2026)
42. Elliott, David. *This happens more frequently than people realize: Arup chief on the lessons learned from a \$25m deepfake crime*. World Economic Forum, Feb 4 2025. Available from <https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/> (cited Aug 15 2026)
43. Anthropic. *Disrupting the first reported AI-orchestrated cyber espionage campaign*. November 2025. Available from <https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf> (cited Aug 15 2026)
44. Internet Watch Foundation. *Full feature-length AI films of child sexual abuse will be 'inevitable' as synthetic videos make huge leaps in sophistication in a year*. July 11, 2025. Available from <https://www.iwf.org.uk/news-media/news/full-feature-length-ai-films-of-child-sexual-abuse-will-be-inevitable-as-synthetic-videos-make-huge-leaps-in-sophistication-in-a-year/> (cited Aug 15 2026); Security Hero, *2023 State of Deepfakes: Realities, Threats, and Impact*, Available from <https://www.securityhero.io/state-of-deepfakes/> (cited Oct 13 2026)
45. Home Office, *Consultation on a new legal framework for law enforcement use of biometrics, facial recognition and similar technologies*, 4 December 2025 – 12 February 2026. Available from <https://www.gov.uk/government/consultations/legal-framework-for-using-facial-recognition-in-law-enforcement/consultation-on-a-new-legal-framework-for-law-enforcement-use-of-biometrics-facial-recognition-and-similar-technologies-accessible>; Webster W., *Commissioner's response to the live facial recognition consultation*. GOV.UK, 2026. Available from <https://www.gov.uk/government/publications/commissioners-response-to-the-live-facial-recognition-consultation/commissioners-response-to-the-live-facial->

- [recognition-consultation-accessible](#) (all cited Aug 15 2026); Big Brother Watch, *Big Brother Watch Responds to the Metropolitan Police's 2025 Live Facial Recognition Report*, Available from <https://bigbrotherwatch.org.uk/press-coverage/big-brother-watch-responds-to-the-metropolitan-polices-2025-live-facial-recognition-report/> (cited Oct 13 2026)
46. Hobbs K.L. & Li B., *Safety, Trust, and Ethics Considerations for Human-AI Teaming in Aerospace Control*, In Computer Science, Nov 15 2023, Available from <https://arxiv.org/abs/2311.08943> (cited Aug 15 2026)
47. Tripp M. et al., *G.M.'s Cruise Moved Fast in the Driverless Race. It Got Ugly*. In *The New York Times*, Nov 3 2023. Available from <https://www.nytimes.com/2023/11/03/technology/cruise-general-motors-self-driving-cars.html>; Kolodny L., *Cruise confirms robotaxis rely on human assistance every 4 to 5 miles*. CNBC, Nov 6 2023. Available from <https://www.cnbc.com/2023/11/06/cruise-confirms-robotaxis-rely-on-human-assistance-every-4-to-5-miles.html>; *Cruise confirms driverless taxis need regular remote help*, In *The Register*, Nov 7 2023. Available from [https://www.theregister.com/2023/11/07/cruise\\_confirms\\_driverless\\_taxis\\_need/](https://www.theregister.com/2023/11/07/cruise_confirms_driverless_taxis_need/) (all cited Aug 15 2026)
48. Zukowski D., *GM Shuts Troubled Cruise Robotaxi Unit*. In Smart Cities Dive, Dec 11 2024. Available from <https://www.smartcitiesdive.com/news/general-motors-shuts-cruise-robotaxi-unit-mary-barra/735205/> (cited Aug 15 2026); NPR, *GM to Retreat from Robotaxis and Stop Funding Its Cruise Autonomous Vehicle Unit*, Dec 11 2024, Available from <https://www.npr.org/2024/12/11/g-s1-37700/gm-to-retreat-from-robotaxis-and-stop-funding-its-cruise-autonomous-vehicle-unit> (cited Oct 13 2026)
49. Ohnsman A., *Argo AI, Ford's Self-Driving Venture With Volkswagen, Is Shutting Down*. In Forbes, Oct 26 2022. Available from <https://www.forbes.com/sites/alanohnsman/2022/10/26/argo-ai-fords-self-driving-venture-with-volkswagen-is-shutting-down/>; *AV developer Argo AI to shut down as Ford, Volkswagen halt investments*. In Smart Cities Dive, Oct 27 2022. Available from <https://www.smartcitiesdive.com/news/av-autonomous-vehicle-argo-ai-shutdown-ford-volkswagen-halt-investment/635089/>; *Ford, VW-backed Argo AI is Shutting Down*. TechCrunch, Oct 26 2022. Available from <https://techcrunch.com/2022/10/26/ford-vw-backed-argo-ai-is-shutting-down/> (all cited Aug 15 2026); The Motley Fool, *Ford (F) Q3 2022 Earnings Call Transcript*, Oct 27 2022, Available from <https://www.fool.com/earnings/call-transcripts/2022/10/27/ford-f-q3-2022-earnings-call-transcript/>; Fortune, *Uber Sells Its Self-Driving Vehicle Unit to Aurora*, Dec 7 2020, Available from <https://fortune.com/2020/12/07/uber-sells-self-driving-vehicle-unit-to-aurora>; SiliconANGLE, *Lyft Selling Self-Driving Car Unit for \$550 Million*, Apr 26 2021, Available from <https://siliconangle.com/2021/04/26/lyft-selling-self-driving-car-unit-550-million/> (all cited Oct 13 2026)

50. Singh S., *The Loop Paradox*, Mar 19 2026, Available from [https://medium.com/@savneetsingh\\_1/the-loop-paradox-human-in-the-loop-human-above-the-loop-ai-in-the-loop-and-human-out-of-the-loop-03fee4d66798](https://medium.com/@savneetsingh_1/the-loop-paradox-human-in-the-loop-human-above-the-loop-ai-in-the-loop-and-human-out-of-the-loop-03fee4d66798) (cited Aug 15 2026)
51. ISACA, *The Growing Challenge of Auditing Agentic AI*, 2025 Available from <https://www.isaca.org/resources/news-and-trends/industry-news/2025/the-growing-challenge-of-auditing-agentic-ai> (cited Aug 15 2026)
52. Raza S. et al., *TRiSM for Agentic AI: A Review of Trust, Risk, and Security Management in LLM-based Agentic Multi-Agent Systems*, June 4 2025, Available from <https://arxiv.org/abs/2506.04133> (cited Aug 15 2026)
53. Cemri M. et al., *Why Do Multi-Agent LLM Systems Fail?*, NeurIPS 2025, Available from <https://arxiv.org/abs/2503.13657> (cited Aug 15 2026)
54. Zou Z. et al., *BlockA2A: Towards Secure and Verifiable Agent-to-Agent Interoperability*, Aug 2 2025, In Computer Science, Available from <https://arxiv.org/abs/2508.01332> (cited Aug 15 2026)
55. Wibowo, Polyzos, *Toward a Safe Internet of Agents*, Nov 29 2025, Available from <https://arxiv.org/abs/2512.00520> (cited Aug 15 2026)
56. McKinsey, *Seizing the Agentic AI Advantage*, June 2025. Available from <https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage> (cited Aug 15 2026)
57. Milei J., *Argentina Invites AI to Free Itself*, In *Financial Times*, June 4 2026, Available from <https://www.ft.com/content/f93022fe-43f7-437d-abd8-06c457c0a43c>; Sircar A., *Argentina Wants to Let AI Own Companies*, In *Forbes*, Jun 10 2026, Available at <https://www.forbes.com/sites/anishasircar/2026/06/10/ai-owned-companies-argentina/> (all cited Aug 15 2026)
58. PYMNTS, *Argentina Drafts Corporate Law that Requires No Human Boss*, Jun 10 2026, Available from <https://www.pymnts.com/news/artificial-intelligence/2026/argentina-drafts-corporate-law-that-requires-no-human-boss/>; Frank Landymore, *Argentina Moves to Legalize Non-Human Corporations Run by AI*, *Futurism*, Jun 9 2026, Available from <https://futurism.com/artificial-intelligence/argentina-legalize-non-human-corporations-ai> (all cited Aug 15 2026)
59. Lynch, A. et al., *Agentic Misalignment: How LLMs Could Be Insider Threats*, Anthropic. June 20 2025. Available from <https://www.anthropic.com/research/agentic-misalignment> (cited Aug 15 2026)
60. Meinke, A. et al., *Frontier Models are Capable of In-Context Scheming*, Apollo Research. Dec 5 2024. Available from <https://www.apolloresearch.ai/science/frontier-models-are-capable-of-incontext-scheming/> (cited Aug 15 2026)
61. Anthropic, *Disrupting the First Reported AI-Orchestrated Cyber Espionage Campaign*, Anthropic Threat Intelligence report, Nov 13 2025. Available from <https://www.anthropic.com/news/disrupting-AI-espionage> (cited Aug 15 2026)

62. OpenAI, *OpenAI and Hugging Face Partner to Address Security Incident During Model Evaluation*, Jul 21 2026 (updated Aug 26 2026), available from <https://openai.com/index/hugging-face-model-evaluation-security-incident/>; OpenAI, *The Hugging Face Incident and the Road Ahead*, Aug 26 2026, available from <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>, with technical report available from <https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf> (all cited Oct 13 2026)
63. Larcher H., Carreira A., G. R., Rannou C., *Anatomy of a Frontier Lab Agent Intrusion: A Technical Timeline of the July 2026 Incident*, Hugging Face, Jul 27 2026, available from <https://huggingface.co/blog/agent-intrusion-technical-timeline> (cited Oct 13 2026)
64. Greenblatt R., Cotra A., Wijk H., *Brief Independent Investigation of Agents' Behavior, Reasoning and Collaboration in the OpenAI / Hugging Face Hacking Incident*, METR, Aug 26 2026, available from <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/> (cited Oct 13 2026)
65. Krakovna V. et al., *Specification Gaming: The Flip Side of AI Ingenuity*, DeepMind, Apr 21 2020, available from <https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/> (cited Oct 13 2026)
66. Von Arx S., Chan L., Barnes E., *Recent Frontier Models Are Reward Hacking*, METR, Jun 5 2025, available from <https://metr.org/blog/2025-06-05-recent-reward-hacking/> (cited Oct 13 2026)
67. UK AI Security Institute, *Cheating Behaviour in Frontier Model Evaluations*, Jul 21 2026, available from <https://www.aisi.gov.uk/blog/cheating-behaviour-in-frontier-model-evaluations> (cited Oct 13 2026)
68. Anthropic, *From Shortcuts to Sabotage: Natural Emergent Misalignment from Reward Hacking*, Nov 21 2025, available from <https://www.anthropic.com/research/emergent-misalignment-reward-hacking> (cited Oct 13 2026)
69. Delimarsky D., *Spec-Driven Development with AI: Get Started with a New Open Source Toolkit*, GitHub Blog, Sept 2 2025, available from <https://github.blog/ai-and-ml/generative-ai/spec-driven-development-with-ai-get-started-with-a-new-open-source-toolkit/> (cited Oct 13 2026)
70. Strathern M., *'Improving Ratings': Audit in the British University System*, European Review 5, no. 3 (1997): 305–321, paraphrasing Goodhart C., *Problems of Monetary Management: The U.K. Experience*, Papers in Monetary Economics, Reserve Bank of Australia, 1975
71. Bondarenko A., Volk D., Volkov D., Ladish J., *Demonstrating Specification Gaming in Reasoning Models*, Palisade Research, Feb 18 2025, available from <https://arxiv.org/abs/2502.13295> (cited Oct 13 2026)

72. Baker B. et al., *Detecting Misbehavior in Frontier Reasoning Models*, OpenAI, Mar 10 2025, available from <https://openai.com/index/chain-of-thought-monitoring/> (cited Oct 13 2026)
73. Zaninotto F., *Spec-Driven Development: The Waterfall Strikes Back*, Marmelab, Nov 12 2025, available from <https://marmelab.com/blog/2025/11/12/spec-driven-development-waterfall-strikes-back.html>; Böckeler B., *Understanding Spec-Driven-Development: Kiro, Spec-Kit, and Tessler*, martinfowler.com, Oct 15 2025, available from <https://martinfowler.com/articles/exploring-gen-ai/sdd-3-tools.html>; for a rebuttal, Brooker M., *Spec Driven Development Isn't Waterfall*, Apr 9 2026, available from <https://brooker.co.za/blog/2026/04/09/waterfall-vs-spec> (all cited Oct 13 2026)
74. NIST, *Technical Blog: Strengthening AI Agent Hijacking Evaluations*, Jan 17 2025, available from <https://www.nist.gov/news-events/news/2025/01/technical-blog-strengthening-ai-agent-hijacking-evaluations> (cited Oct 13 2026)
75. Anthropic, *Improve Your Prompts in the Developer Console*, Oct 14 2024, available from <https://claude.com/resources/articles/prompt-improver>; Swaminathan N. & Singh D., *Introducing Kiro*, Jul 14 2025, available from <https://kiro.dev/blog/introducing-kiro/> (both cited Oct 13 2026)
76. Bainbridge L. *Ironies of automation*. In *Automatica* 19, no. 6 (November 1983): 775–779. Available from [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8) (cited Aug 15 2026)
77. Bengio Y. et al., *Behavioural Red Lines as a Safety Framework for Intolerable Levels of Risk*. In *The Artificial Intelligence Revolution*, Springer Nature, 2026 (cited Aug 15 2026)
78. Pan et al., *Frontier AI Systems Have Surpassed the Self-Replicating Red Line*, Dec 9 2024, Available from <https://arxiv.org/pdf/2412.12140> (cited Aug 15 2026)
79. Yudkowsky E. et al., *If Anyone Builds It, Everyone Dies: The Case Against Superintelligent AI*. London: The Bodley Head, 2025. ISBN 978-1-847-92892-4.
80. *Letter from OpenAI whistleblower(s)' counsel to Gary Gensler, Chair, U.S. Securities and Exchange Commission*, 1 July 2024. Reference: *OpenAI Violations of Rule 21F-17(a) and Implementation of E.O. 14110*. Available from <https://www.washingtonpost.com/documents/83df0e55-546c-498a-9efc-06fac591904e.pdf> (cited Aug 15 2026)
81. Mitchell M. et al., *Model Cards for Model Reporting*, In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19)*, 220–229. New York: Association for Computing Machinery, 2019. Available from <https://doi.org/10.1145/3287560.3287596> (cited Aug 15 2026)
82. Gartner, *Is Synthetic Data the Future of AI?*, press release, Jun 22 2022, Available from <https://www.gartner.com/en/newsroom/press-releases/2022-06-22-is-synthetic-data-the-future-of-ai> (cited Oct 13 2026)

83. Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. *AI models collapse when trained on recursively generated data*. In *Nature* 631, no. 8022 (2024): 755–759. Available from <https://www.nature.com/articles/s41586-024-07566-y> (cited Aug 15 2026)
84. Decruyenaere A. et al., *The Real Deal Behind the Artificial Appeal: Inferential Utility of Tabular Synthetic Data*. In Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence, PMLR 244 (2024): 966–996. Available from <https://arxiv.org/abs/2312.07837>; Alaa, A. M. et al., *How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Models*. In Proceedings of the 39th International Conference on Machine Learning, PMLR 162 (2022): 290–306. Available from <https://arxiv.org/abs/2102.08921>; van Breugel B. et al., *Membership Inference Attacks against Synthetic Data through Overfitting Detection*. In Proceedings of the 26th International Conference on Artificial Intelligence and Statistics, PMLR 206 (2023): 3493–3514. Available from <https://arxiv.org/abs/2302.12580> (all cited Aug 15 2026)
85. Stadler T. et al., *Synthetic Data – Anonymisation Groundhog Day*, 31st USENIX Security Symposium, Available from <https://www.usenix.org/conference/usenixsecurity22/presentation/stadler> (cited Aug 15 2026)
86. Bender, E. M. et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. New York: Association for Computing Machinery, 2021. Available from <https://doi.org/10.1145/3442188.3445922> (cited Aug 15 2026)
87. Georgieva K., *AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity*. IMF Blog, 14 January 2024. Available from <https://www.imf.org/en/blogs/articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>; Underlying paper: Cazzaniga M. et al., *Gen-AI: Artificial Intelligence and the Future of Work*. IMF Staff Discussion Note SDN/2024/001, January 2024. Available from <https://www.imf.org/-/media/files/publications/sdn/2024/english/sdnea2024001.pdf> (cited Aug 15 2026)
88. Briggs J. et al., *The Potentially Large Effects of Artificial Intelligence on Economic Growth*. Goldman Sachs Global Investment Research, Global Economics Analyst, 26 March 2023. Available from <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent> (cited Aug 15 2026)
89. McKinsey Global Institute. *The Economic Potential of Generative AI: The Next Productivity Frontier*. June 2023. Available from <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (cited Aug 15 2026)

90. Acemoglu D., *The Simple Macroeconomics of AI*. In *Economic Policy* 40, no. 121 (January 2025): 13–58. Available from <https://doi.org/10.1093/epolic/eiae042> (cited Aug 15 2026)
91. The Economist, *Meet the World's Top AI-pilled Economists*, June 15 2026, Available from <https://www.economist.com/finance-and-economics/2026/06/15/meet-the-worlds-top-ai-pilled-economists> (cited Aug 15 2026)
92. Wakefield J., *Google's Ethics Board Shut Down*, Apr 5 2019, Available from <https://www.bbc.com/news/technology-47825833> (cited Aug 15 2026)
93. Washington Post, *Google Hired Timnit Gebru To Be an Outspoken Critic of Unethical AI. Then She Was Fired For It*. Dec 23 2020. Available from <https://www.washingtonpost.com/technology/2020/12/23/google-timnit-gebru-ai-ethics/> (cited Aug 15 2026); Axios, *Google Fires Another AI Ethics Leader*, Feb 19 2021, Available from <https://www.axios.com/2021/02/19/google-fires-another-ai-ethics-leader> (cited Oct 13 2026)
94. *EU Guidelines: Ethics Washing Made in Europe*. Available from [https://static1.squarespace.com/static/627b272057953023992b01d0/t/64675706e699622b2d41002b/1684494086647/Metzinger+\(2019\)+EU+Guidelines\\_Ethics+Washing+Made+in+Europe.pdf](https://static1.squarespace.com/static/627b272057953023992b01d0/t/64675706e699622b2d41002b/1684494086647/Metzinger+(2019)+EU+Guidelines_Ethics+Washing+Made+in+Europe.pdf) (cited Aug 15 2026)
95. Nvidia. *Nvidia CEO: Every Country Needs AI*. Available from <https://blogs.nvidia.com/blog/world-governments-summit/> (cited Aug 15 2026); The Economist, *Can Nvidia Persuade Governments To Pay For Sovereign AI?* Jul 13 2025, available from <https://www.economist.com/business/2025/07/13/can-nvidia-persuade-governments-to-pay-for-sovereign-ai> (cited Aug 15 2026)
96. *Interview between Alex Karp and Axel Springer CEO Mathias Döpfner*, MDMEETS, Jul 13 2026. Available from <https://www.youtube.com/watch?v=xm0rGvvxVOM> (cited Aug 15 2026)
97. The Guardian, Booth R., *MPs Urge Labour to Ditch 330m Palantir Software Contract with NHS*. Jul 9 2026. Available from <https://www.theguardian.com/politics/2026/jul/09/mps-urge-labour-to-ditch-330m-palantir-software-contract-with-nhs> (cited Aug 15 2026); The Register, *UK Lawmakers Call on Government to Ditch Palantir NHS Contract*, Jun 3 2026, Available from <https://www.theregister.com/software/2026/06/03/uk-lawmakers-call-on-government-to-ditch-palantir-nhs-contract/5250150> (cited Oct 13 2026)
98. Lauterbach A., *Unitarism vs. Individuality and a New Digital Agenda: The Power of Decentralized Web*. In *Frontiers in Human Dynamics*, 2021. Available from <https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2021.626299/full> (cited Aug 15 2026)
99. <https://www.midata.coop/en/home/>
100. CGI, Schippert R., *Unlocking Collaborative Data Ecosystems with Catena-X*, Mar 4 2025, Available from <https://www.cgi.com/en/blog/manufacturing/unlocking-collaborative-data->

- [ecosystems-catena-x](#) (cited Aug 15 2026); DLR, *Catena-X*, project page, Available from <https://www.dlr.de/en/ki/research-transfer/projects/catena-x> (cited Oct 13 2026)
101. <https://oceanprotocol.com/> (cited Aug 15 2026)
  102. Zurich Insurance Group, *Zurich to Launch New Apprenticeship Program in U.S.*, press release, Jun 9 2015, Available from <https://www.zurich.com/media/news-releases/2015/2015-0609-01> (cited Aug 15 2026); Insurance Business, *Zurich Unlocking Talent Pool with Expanded Apprenticeship Program*, Jan 7 2022, Available from <https://www.insurancebusinessmag.com/us/news/breaking-news/zurich-unlocking-talent-pool-with-expanded-apprenticeship-program-321360.aspx>; Digital Insurance, *Zurich Announces New Bachelor's Degree Apprenticeship Program*, Sept 19 2022, Available from <https://dig-in.com/news/zurich-announces-new-bachelors-degree-apprenticeship-program> (both cited Oct 13 2026)
  103. Siemens, *Siemens Sets New Training Standards for GenZ*, press release, Sept 3 2025, Available from <https://press.siemens.com/global/en/node/7213>; BIBB AusbildungPlus, *Künstliche Intelligenz (B.Sc.), dualer Studiengang*, OTH Amberg-Weiden with Siemens, Available from [https://www.bibb.de/dienst/ausbildungplus/de/dualer\\_studiengang/ansehen/20426](https://www.bibb.de/dienst/ausbildungplus/de/dualer_studiengang/ansehen/20426) (both cited Oct 13 2026)
  104. OpenAI, *Introducing GPT-6.1 Sol*, Sept 29 2026, available from <https://openai.com/index/introducing-gpt-6-1-sol/>; OpenAI, *Better Prompt Caching for GPT-6*, Sept 22 2026, available from <https://openai.com/index/better-prompt-caching-for-gpt-6/>; OpenAI, *DevDay 2026 Recap*, Sept 29 2026, available from <https://openai.com/index/devday-2026-recap/>; OpenAI, *Using GPT-6*, API documentation, available from <https://developers.openai.com/api/docs/guides/latest-model> (all cited Oct 13 2026)
  105. Casado M. et al., *The New Business of AI (and How It's Different From Traditional Software)*, February 16 2020, Available from <https://a16z.com/the-new-business-of-ai-and-how-its-different-from-traditional-software/> (cited Aug 15 2026); Dmitry Fisher talked about the optimistic side of Casado's and Bornstein's assessment in Episodes 23 (*All Senses on Deck: How Multimodal AI is Changing the Game with Dr. Dmitry Fisher*) and 28, *Vision Quest: The Future Through AI's Eyes with Dr. Dmitry Fisher*)
  106. Mellon Investments, *Record-Breaking AI-Related Debt Issuance in 2025*, December 2025 (roughly \$121 billion in new debt in 2025, with over \$90 billion raised in the final three months), Available from <https://www.mellon.com/insights/insights-articles/record-breaking-ai-related-debt-issuance-in-2025.html>; MarketScreener, *JPMorgan Chase Analysts See \$5.5 Trillion in AI Buildout Costs by 2030*, Available from <https://www.marketscreener.com/news/jpmorgan-chase-analysts-see-5-5->

- [trillion-in-ai-buildout-costs-by-2030-bloomberg-reports-ce7f5cdfde8ff027](#) (both cited Aug 15 2026); Bloomberg Law, *JPMorgan Sees AI Buildout Costs Totaling \$5.5 Trillion by 2030*, Available from <https://news.bloomberglaw.com/texas-brief/jpmorgan-sees-ai-buildout-costs-totaling-5-5-trillion-by-2030> (cited Oct 13 2026)
107. Bisnow, *Meta Pushes Its Largest Data Center Project Off Its Books With \$27B Joint Venture*, Oct 22 2025, available from <https://www.bisnow.com/news/national/data-center-capital-markets/meta-pushes-its-largest-data-center-project-off-its-books-with-27b-joint-venture-131490>; Bank of England, *Financial Stability Report - December 2025*, Dec 2 2025, available from <https://www.bankofengland.co.uk/financial-stability-report/2025/december-2025> (both cited Oct 13 2026)
108. Yahoo Finance/Quartz, citing Bloomberg, on xAI's special purpose vehicle for Nvidia chips, Oct 8 2025, available from <https://finance.yahoo.com/news/elon-musk-xai-raise-20-124933356.html>; DatacenterDynamics, on the \$38bn debt package for Oracle-linked data centers, Sept 5 2025, available from <https://datacenterdynamics.com/en/news/jpmorgan-chase-and-mitsubishi-ufj-lead-38bn-debt-package-for-oracle-linked-data-centers-report/> (both cited Oct 13 2026)
109. NVIDIA, *OpenAI and NVIDIA Announce Strategic Partnership to Deploy 10GW of NVIDIA Systems*, Sept 22 2025, available from <https://nvidianews.nvidia.com/news/openai-and-nvidia-announce-strategic-partnership-to-deploy-10gw-of-nvidia-systems>; AMD, *AMD and OpenAI Announce Strategic Partnership*, Oct 6 2025, available from <https://newsroom.amd.com/news/amd-and-openai-announce-strategic-partnership-to-d/> (both cited Oct 13 2026)
110. Fortune, on Moody's assessment of \$662 billion in off-balance-sheet data center commitments, Feb 25 2026, available from <https://fortune.com/2026/02/25/hyperscaler-risk-off-balance-sheet-662-billion-data-center-commitments-meta-amazon-microsoft-oracle-alphabet/>; IFR, *Moody's Opinion Threatens to Derail Off-Balance-Sheet Data Centre Deals*, Feb 23 2026, available from <https://www.ifre.com/bonds/2390941/moodys-opinion-threatens-to-derail-off-balance-sheet-data-centre-deals> (both cited Oct 13 2026)
111. BIS, *Financing the AI Infrastructure Boom: On- and Off-Balance Sheet Borrowing*, BIS Quarterly Review, Mar 16 2026, available from [https://www.bis.org/publ/qtrpdf/r\\_qt2603u.htm](https://www.bis.org/publ/qtrpdf/r_qt2603u.htm); BIS, *Annual Economic Report 2026*, Chapter I: *Progress and Peril*, Jun 28 2026, available from <https://www.bis.org/publications/aer-2026/progress-peril> (both cited Oct 13 2026)
112. Bank of England, *Financial Stability Report - July 2026*, Jul 7 2026, available from <https://www.bankofengland.co.uk/financial-stability-report/2026/july-2026> (cited Oct 13 2026)

113. InvestmentNews, *IMF Warns Bond Yields Face Disorderly Spike on Debt and AI Risks*, Oct 13 2026, reporting on the IMF Global Financial Stability Report of October 2026, available from <https://www.investmentnews.com/alternatives/imf-warns-bond-yields-face-disorderly-spike-on-debt-and-ai-risks/268570>; IMF, *Global Financial Stability Report*, April 2026, Chapter 1, available from <https://www.imf.org/-/media/files/publications/gfsr/2026/april/english/ch1.pdf> (both cited Oct 13 2026)
114. Friedman D., *The \$176 Billion Accounting Question*, Nov 2025, summarising Michael Burry's critique of GPU depreciation periods, available from <https://davefriedman.substack.com/p/the-176-billion-accounting-question> (cited Oct 13 2026)
115. Anthropic Engineering, *Demystifying Evals for AI Agents*, Jan 9 2026, available from <https://www.anthropic.com/engineering/demystifying-evals-for-ai-agents> (defines the agent harness or scaffold as “the system that enables a model to act as an agent: it processes inputs, orchestrates tool calls, and returns results”); Anthropic Engineering, *Effective Harnesses for Long-Running Agents*, Nov 26 2025, available from <https://www.anthropic.com/engineering/effective-harnesses-for-long-running-agents>; OpenAI, *Unrolling the Codex Agent Loop*, Jan 23 2026, available from <https://openai.com/index/unrolling-the-codex-agent-loop/> (all cited Oct 13 2026)
116. OpenAI, *Developers Changelog* (Agents API public beta, Sept 10 2026), available from <https://developers.openai.com/changelog> (cited Oct 13 2026)
117. OpenAI, *Reasoning Models*, API documentation, available from <https://developers.openai.com/api/docs/guides/reasoning>; Anthropic, *Effort*, Claude documentation, available from <https://platform.claude.com/docs/en/build-with-claude/effort>; Google, *Gemini Thinking*, Gemini API documentation, available from <https://ai.google.dev/gemini-api/docs/thinking> (all cited Oct 13 2026). All three bill reasoning tokens as output tokens; effort is described as guidance on how much to think, not a strict token budget. On current OpenAI (GPT-5.6 and later) and Anthropic models, earlier reasoning is kept in the conversation by default and billed as input in later requests.
118. Artificial Analysis, *Intelligence Benchmarking Methodology*, available from <https://artificialanalysis.ai/methodology/intelligence-benchmarking>; Artificial Analysis, *GPT-6.1 Sol Replaces GPT-6 Sol After Just 7 Days, with Near-Astra Intelligence*, Sept 29 2026, available from <https://artificialanalysis.ai/articles/gpt-6-1-sol-replaces-gpt-6-sol-after-just-7-days-with-near-astra-intelligence>; Artificial Analysis, *Claude Fable 5.1*, Sept 1 2026, available from <https://artificialanalysis.ai/articles/claude-fable-5-1>; Artificial Analysis, *Recent Open Weights Model Launches*, Apr 2026, available from <https://artificialanalysis.ai/articles/recent-open-weights-model-launches>;

Artificial Analysis, *LLM Leaderboard and Open Weights Models*, available from <https://artificialanalysis.ai/leaderboards/models> and <https://artificialanalysis.ai/models/open-source> (all cited Oct 13 2026). Note: index scores from different index versions are not directly comparable.

119. Anthropic, *Manage Costs Effectively*, Claude Code documentation, available from <https://code.claude.com/docs/en/costs> ("Claude Code sends your full conversation with every request [...] Claude Code re-reads that history at the cached token rate"); Bai et al., arXiv:2604.22750, 2026, available from <https://arxiv.org/html/2604.22750v3> ("cache reads dominate both raw token volume and dollar cost") (both cited Oct 13 2026)
120. Menlo Ventures, *2025: The State of Generative AI in the Enterprise*, available from <https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/>; Wang S. et al., *Leaders, Gainers and Unexpected Winners in the Enterprise AI Arms Race*, a16z, Jan 30 2026, available from <https://a16z.com/leaders-gainers-and-unexpected-winners-in-the-enterprise-ai-arms-race/> (both cited Aug 15 2026)
121. Lauterbach A., *AI Literacy in SMEs*. Original Survey, n=915 (people who opened the survey), status October 2026; questions Q21-Q23 on foundational models (percentages refer to respondents answering each question); (cited Oct 13 2026)
122. Epoch AI, *Data Insight: US vs. China Frontier AI Capabilities (Epoch Capabilities Index)*, Jan 2 2026, available from <https://epoch.ai/data-insights/us-vs-china-eci> (cited Oct 13 2026)
123. NIST, *CAISI Evaluation of DeepSeek V4 Pro*, May 1 2026, available from <https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro> (cited Oct 13 2026)
124. NIST, *Evaluation of DeepSeek AI Models*. Available from [https://www.nist.gov/system/files/documents/2025/09/30/CAISI\\_Evaluation\\_of\\_DeepSeek\\_AI\\_Models.pdf](https://www.nist.gov/system/files/documents/2025/09/30/CAISI_Evaluation_of_DeepSeek_AI_Models.pdf) (cited Aug 15 2026)
125. ChinaTalk, translation of the 36Kr interview with DeepSeek founder Liang Wenfeng, July 2024, available from <https://www.chinatalk.media/p/deepseek-ceo-interview-with-chinas>; Chow V., *China AI Has Less than 20% Chance to Exceed US over Next 3-5 Years: Alibaba Scientist*, South China Morning Post, Jan 12 2026, available from <https://www.scmp.com/tech/big-tech/article/3339527/china-ai-has-less-20-chance-exceed-us-over-next-3-5-years-alibaba-scientist>; Shen X., *Alibaba Chairman Joe Tsai Says US-China AI Competition Not Winner-Takes-All*, South China Morning Post, Oct 8 2025, available from <https://www.scmp.com/tech/big-tech/article/3328272/alibaba-chairman-joe-tsai-says-us-china-ai-competition-not-winner-takes-all>; State Council of the PRC, *Opinions on Deepening the Implementation of the "AI+" Action*, Aug 2025, translation by CSET, available from [https://cset.georgetown.edu/wp-content/uploads/t0652\\_AI\\_plus\\_opinions\\_EN.pdf](https://cset.georgetown.edu/wp-content/uploads/t0652_AI_plus_opinions_EN.pdf) (all cited Oct 13 2026)

126. Artificial Analysis, *Text-to-Video*, *Image-to-Video* and *Text-to-Image* leaderboards, available from <https://artificialanalysis.ai/video/leaderboard/text-to-video>, <https://artificialanalysis.ai/video/leaderboard/image-to-video> and <https://artificialanalysis.ai/image/leaderboard/text-to-image> (all accessed Oct 13 2026)
127. OpenAI, *Whisper* (code and model weights under the MIT License), available from <https://github.com/openai/whisper>; OpenAI, *Introducing gpt-oss*, Aug 5 2025, available from <https://openai.com/index/introducing-gpt-oss/>; Google, *Gemma 4*, Apr 2 2026, available from <https://blog.google/innovation-and-ai/technology/developers-tools/gemma-4/>; Mistral AI, *Introducing Mistral 3*, Dec 2 2025, available from <https://mistral.ai/news/mistral-3> (all cited Oct 13 2026)
128. Orange, press release on deploying OpenAI's gpt-oss models on its own infrastructure, Aug 5 2025, available from <https://newsroom.orange.com/?p=58796> (cited Oct 13 2026)
129. OpenAI, announcement on expanding access to ChatGPT (more than one billion weekly active users), Aug 31 2026, available from <https://openai.com/index/expanding-access-to-ai-with-chatgpt-ads/>; Google, announcement on the Gemini app reaching more than one billion monthly users, Aug 11 2026, available from <https://blog.google/innovation-and-ai/products/gemini-app/one-billion-monthly-users/> (both cited Oct 13 2026)
130. Eurostat, news release on the use of generative AI tools by individuals in the EU in 2025, Dec 16 2025, available from <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20251216-3> (cited Oct 13 2026)
131. Chatterji A., Cunningham T., Deming D., Hitzig Z., Ong C., Shan C.Y., Wadman K., *How People Use ChatGPT*, NBER Working Paper 34255, Sept 2025, available from <https://www.nber.org/papers/w34255> (cited Oct 13 2026)
132. Anthropic, *Anthropic Economic Index Report*, January 2026, available from <https://www.anthropic.com/research/anthropic-economic-index-january-2026-report>; Anthropic, *Anthropic Economic Index Report*, March 2026, available from <https://www.anthropic.com/research/economic-index-march-2026-report>; Anthropic, *Anthropic Raises \$30 Billion Series G Funding at \$380 Billion Post-Money Valuation*, Feb 2026, available from <https://www.anthropic.com/news/anthropic-raises-30-billion-series-g-funding-380-billion-post-money-valuation> (all cited Oct 13 2026)
133. The Decoder, *OpenAI Claims a Breakthrough in LLM Reasoning on Complex Math Problems*, Jul 2025, available from <https://the-decoder.com/openai-claims-a-breakthrough-in-llm-reasoning-on-complex-math-problems/>; OpenAI, *Accelerating Science with GPT-5*, available from <https://openai.com/index/accelerating-science-gpt-5/>; OpenAI, *Model Disproves Discrete Geometry Conjecture*, May 2026, available from <https://openai.com/index/model-disproves-discrete-geometry-conjecture/> (all cited Oct 13 2026)

134. Pichai S., keynote remarks at Google I/O 2026, available from <https://blog.google/innovation-and-ai/sundar-pichai-io-2026/>; Google DeepMind, *Advanced Version of Gemini with Deep Think Officially Achieves Gold-Medal Standard at the International Mathematical Olympiad*, Jul 21 2025, available from <https://deepmind.google/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/> (both cited Oct 13 2026)
135. DeepSeek, *DeepSeek-V4-Pro* model card (MIT License), available from <https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro>; Moonshot AI, *Kimi-K3* model card (Kimi K3 License), available from <https://huggingface.co/moonshotai/Kimi-K3>; Lardinois F., *Z.ai's GLM-5.3 Goes Open Weight, but Its New License Aims at Hyperscalers*, The New Stack, Aug 28 2026, available from <https://thenewstack.io/zai-glm-weights-license/> (all cited Oct 13 2026)
136. Wharton. Human-AI Research. *Accountable Acceleration: GenAI Fast-Tracks Into the Enterprise*. Oct 2025. Available from [https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report\\_Full-Report.pdf](https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report_Full-Report.pdf) (cited Aug 15 2026)
137. SBE Council, *Small Business Technology Use Survey*, March 2026 (firms with 2-99 employees), available from <https://sbecouncil.org/wp-content/uploads/2026/03/SBE-Technology-Use-Survey-March-2026-Final-2.pdf> (cited Oct 13 2026)
138. Datasumi, *Europe Steps Up AI Pace, But Small Businesses Lag Behind: Why the Gap Widens*. July 15 2026. Available from <https://www.datasumi.com/blog/eu-ai-small-business-adoption-gap> (cited Aug 15 2026)
139. Anthropic, *Claude for Nonprofits*. Dec 2 2025. Available from <https://www.anthropic.com/news/claude-for-nonprofits> (cited Aug 15 2026)
140. Google Support, *Workspace for Nonprofits AI & Productivity Program*. Available from <https://support.google.com/nonprofits/answer/17049066?hl=en> (cited Aug 15 2026)
141. GSA, *Buy AI*, Available from <https://www.gsa.gov/artificial-intelligence/buy-ai> (cited Aug 15 2026); Mobile World Live, *OpenAI Swaps \$1 Government Pricing for 50% Discount*, Available from <https://mobileworldlive.com/ai-cloud/openai-swaps-1-government-pricing-for-50-discount> (cited Oct 13 2026)
142. Garland M., *The Coming AI Reckoning: Slouching Toward Vendor Lock*, Federal News Network, June 8 2026, Available from <https://federalnewsnetwork.com/commentary/2026/06/the-coming-ai-reckoning-slouching-toward-vendor-lock/> (cited Aug 15 2026)
143. TechSoup, *Microsoft to Discontinue Select Donated Microsoft 365 and Office 365 Plans*, May 15 2025. Available from <https://blog.techsoup.org/en-us/posts/microsoft-to-discontinue-select-donated-microsoft-365-and-office-365-plans> (cited Aug 15 2026); Whole Whale, *400K Nonprofits Impacted by*

- Microsoft Grant Cancellation Coming July 2025*, Available from <https://wholewhale.com/tips/400k-nonprofits-impacted-by-microsoft-grant-cancellation-coming-july-2025/> (cited Oct 13 2026)
144. Saran C., *Education Sector Facing Huge VMware Cost Increases After Broadcom Ends Discounts*, Computer Weekly, Apr 4 2024, Available from <https://www.computerweekly.com/news/366579593/Education-sector-facing-huge-VMware-cost-increases-after-Broadcom-ends-discounts> (cited Aug 15 2026)
  145. EpochAI, *The US Hosts the majority of GPU Cluster Performance, Followed by China*. Available from <https://epoch.ai/data-insights/ai-supercomputers-performance-share-by-country>; GeoCoded, *State of Global AI Compute*. 2025 Edition. Available from <https://www.sanchez.vc/geocoded-special-reports/state-of-global-ai-compute-2025-edition>; Sangalli I., Nasdaq 2025, *Beyond the Headlines – the AI Race Explained: US vs China*, Available from <https://www.nasdaq.com/articles/global-indexes/beyond-the-headlines-the-ai-race-explained-us-vs-china> (cited Aug 15 2026)
  146. Phillips-Robins A. et al., *The Compute Coalition: How to Build the Future of AI in the Free World*, Jun 8 2026, Available from <https://carnegieendowment.org/research/2026/06/the-compute-coalition-how-to-build-the-future-of-ai-in-the-free-world> (cited Aug 15 2026)
  147. Fried I., *Scoop: Meta Won't Offer Future Multimodal AI Models in EU*, in Axios, Available from <https://www.axios.com/2024/07/17/meta-future-multimodal-ai-models-eu> (cited Aug 15 2026)
  148. Meta, *Llama 4 Acceptable Use Policy*, available from <https://dev.meta.ai/llama/llama4/use-policy> (cited Oct 13 2026)
  149. Garante per la protezione dei dati personali, order against Hangzhou DeepSeek and Beijing DeepSeek, Jan 30 2025, available from <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10098477>; Berlin Commissioner for Data Protection and Freedom of Information, press release on DeepSeek, Jun 27 2025, available from [https://www.datenschutz-berlin.de/fileadmin/user\\_upload/pdf/pressemitteilungen/2025/20250627-BlnBDI-Press-Release\\_DeepSeek.pdf](https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/pressemitteilungen/2025/20250627-BlnBDI-Press-Release_DeepSeek.pdf) (both cited Oct 13 2026)
  150. Bitkom, *KI aus China für deutsche Wirtschaft kein Thema*, Sept 9 2026, telephone survey of 603 companies with 20 or more employees, available from <https://www.bitkom.org/Presse/Presseinformation/KI-aus-China-fuer-deutsche-Wirtschaft-kein-Thema> (cited Oct 13 2026)
  151. Hugging Face, *Pickle Scanning*, documentation, available from <https://huggingface.co/docs/hub/security-pickle>; Hubinger E. et al., *Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training*, Anthropic, Jan 14 2024, available from <https://arxiv.org/abs/2401.05566>; CrowdStrike, *CrowdStrike Researchers Identify Hidden Vulnerabilities in AI-Coded Software*, Nov 20 2025, available from

- <https://www.crowdstrike.com/en-us/blog/crowdstrike-researchers-identify-hidden-vulnerabilities-ai-coded-software/> (all cited Oct 13 2026)
152. Amazon Web Services, *Amazon Bedrock Pricing*, available from <https://aws.amazon.com/bedrock/pricing/>; Amazon Bedrock model cards (regional availability), e.g. *DeepSeek-V3.2*, available from <https://docs.aws.amazon.com/bedrock/latest/userguide/model-card-deepseek-deepseek-v3-2.md> (both cited Oct 13 2026)
153. Microsoft, *Region Availability for Foundry Models Sold by Azure*, available from <https://learn.microsoft.com/en-us/azure/foundry/foundry-models/concepts/models-sold-directly-by-azure-region-availability?tabs=az-america&pivots=standard>; Microsoft, *Deployment Types for Microsoft Foundry Models*, available from <https://learn.microsoft.com/en-us/azure/foundry/foundry-models/concepts/deployment-types> (both cited Oct 13 2026)
154. ETH Zurich, *Apertus: A Fully Open, Transparent, Multi-Lingual Language Model*. Sept 2 2025. Available from <https://ethz.ch/en/news-and-events/eth-news/news/2025/09/press-release-apertus-a-fully-open-transparent-multilingual-language-model.html> (cited Aug 15 2026)
155. Laurer M., Renda A., Yeung T., *Clarifying the Costs for the EU's AI Act*, CEPS, Sept 24 2021, Available from <https://www.ceps.eu/?p=34112> (cited Oct 13 2026)
156. Deloitte, *The Path to Sustainable Generative AI Value Balances Passion, Pragmatism and Patience, Finds New Deloitte Survey*. Jan 21 2025. Available from <https://www.deloitte.com/us/en/about/press-room/state-of-generative-ai.html> (cited Aug 15 2026)
157. European Commission, *Employment and Social Developments in Europe 2023*, ch. 2.4. Available from <https://op.europa.eu/webpub/empl/esde-2023/chapters/chapter-2-4-1.html> (cited Aug 15 2026)
158. European Commission, *2024 Ageing Report: Economic and Budgetary Projections for the EU Member States (2022–2070)*, Institutional Paper 279. Available from [https://economy-finance.ec.europa.eu/publications/2024-ageing-report-economic-and-budgetary-projections-eu-member-states-2022-2070\\_en](https://economy-finance.ec.europa.eu/publications/2024-ageing-report-economic-and-budgetary-projections-eu-member-states-2022-2070_en) (cited Aug 15 2026)
159. Pinkus D. & Ruer N., *The Demographic Divide: Inequalities in Ageing Across the European Union*, Bruegel Policy Brief 13/2025, Mar 27 2025. Available from <https://www.bruegel.org/policy-brief/demographic-divide-inequalities-ageing-across-european-union> (cited Aug 15 2026)
160. OECD, *OECD Employment Outlook 2025: Job Markets Remain Resilient but Population Ageing Will Cause Significant Labour Shortages and Fiscal Pressures*, press release, Jul 9 2025. Available from <https://www.oecd.org/en/about/news/press-releases/2025/07/oecd-job-markets-remain-resilient-but-population-ageing-will-cause-significant-labour-shortages-and-fiscal-pressures.html> (cited Oct 13 2026)

161. Bodnár K. & Nerlich C., *The Macroeconomic and Fiscal Impact of Population Ageing*, ECB Occasional Paper 296, June 2022. Available from <https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op296~aaf209ffe5.en.pdf> (cited Aug 15 2026)
162. Commission Decision 2000/518/ EC of July 26 2000 on the adequate protection of personal data provided in Switzerland, Official Journal L 215, Aug 25 2000. Available from <https://eur-lex.europa.eu/eli/dec/2000/518/oj/eng> (cited Aug 15 2026)
163. Department for Science, Innovation and Technology (DSIT), *AI Growth Lab: A Call for Evidence*, issued Oct 21 2025, closed Jan 7 2026. Available from <https://www.gov.uk/government/calls-for-evidence/ai-growth-lab> (cited Aug 15 2026); GOV.UK, *New Blueprint for AI Regulation Could Speed Up Planning Approvals, Slash NHS Waiting Times and Drive Growth and Public Trust*, Oct 21 2025, Available from <https://www.gov.uk/government/news/new-blueprint-for-ai-regulation-could-speed-up-planning-approvals-slash-nhs-waiting-times-and-drive-growth-and-public-trust>; UKAuthority, *DSIT Scrapped as Burnham Government Reshapes Whitehall Tech Functions*, Available from <https://www.ukauthority.com/articles/dsit-scrapped-as-burnham-government-reshapes-whitehall-tech-functions> (both cited Oct 13 2026)
164. IAPP, *King's Speech Signals Diffuse UK Digital Policy Agenda, but No AI Bill*, May 2026, available from <https://iapp.org/news/a/king-s-speech-signals-diffuse-uk-digital-policy-agenda-but-no-ai-bill>; UK Parliament, Written Statement HLWS298, *Machinery of Government Changes*, Jul 21 2026, available from <https://questions-statements.parliament.uk/written-statements/detail/2026-07-21/hlws298> (both cited Oct 13 2026)
165. Ollama, *FAQ and Hardware Support*, documentation, available from <https://docs.ollama.com/faq> and <https://docs.ollama.com/gpu>; Apple, *Mac Studio - Technical Specifications*, available from <https://www.apple.com/mac-studio/specs/>; NVIDIA, *DGX Spark*, available from <https://www.nvidia.com/en-us/products/workstations/dgx-spark/>; LearnWithHasan, *Self-Host Ollama* (CPU and GPU benchmarks), Aug 12 2026, available from <https://learnwithhasan.com/guide/self-host-ollama/> (all cited Oct 13 2026)
166. Hugging Face, *Optimizing LLMs for Speed and Memory*, Transformers documentation, available from [https://huggingface.co/docs/transformers/llm\\_tutorial\\_optimization](https://huggingface.co/docs/transformers/llm_tutorial_optimization); llama.cpp, *Quantize* (bits per weight by quantization type), available from <https://github.com/ggml-org/llama.cpp/blob/master/tools/quantize/README.md> (both cited Oct 13 2026)
167. Ollama, model library: *Llama 3.1* and *Gemma 4* tags, available from <https://ollama.com/library/llama3.1/tags> and <https://ollama.com/library/gemma4/tags>; Ollama, *Context Length*, documentation, available from <https://docs.ollama.com/context-length> (all cited Oct 13 2026)

168. Ollama, available from <https://ollama.com>; LM Studio, available from <https://lmstudio.ai>; vLLM documentation, available from <https://docs.vllm.ai>; NVIDIA, GeForce RTX 5090, available from <https://www.nvidia.com/en-us/geforce/graphics-cards/50-series/rtx-5090/> (all cited Oct 13 2026)
169. TensorRigs, *How Much RAM for LocalLLMs? The Complete 2026 Guide*. Mar 26 2026. Available from <https://tensorrigs.com/blog/ram-for-local-llm/>; OverchatAI, *Local LLM Hardware Requirements in 2026*. Available from <https://overchat.ai/ai-hub/llm-hardware-requirements> (both cited Aug 15 2026)
170. Roche, *Roche Launches NVIDIA AI Factory to Accelerate the Development of New Therapeutics and Diagnostics Solutions*. March 16 2026. Available from <https://www.roche.com/media/releases/med-cor-2026-03-16> (cited Aug 15 2026)
171. Lilly, *Lilly Partners with Nvidia to Build the Industry's Most Powerful AI Supercomputer, Supercharging Medicine Discovery and Delivery for Patients*. Oct 28 2025. Available from <https://investor.lilly.com/news-releases/news-release-details/lilly-partners-nvidia-build-industrys-most-powerful-ai> (cited Aug 15 2026)
172. R&D World, *Lilly and Nvidia's AI Push: A South San Francisco Lab and Pharma's Biggest Supercomputer Take Aim at Drug Discovery's Hard Limits*, Jan 2026, available from <https://www.rdworldonline.com/lilly-and-nvidias-ai-push-a-south-san-francisco-lab-and-pharmas-biggest-supercomputer-take-aim-at-drug-discoverys-hard-limits/>; Fierce Pharma, *Lilly Debuts Nvidia Supercomputer with Fanfare and Focus on Escaping Traditional Pharma Lifecycle*, available from <https://www.fiercepharma.com/biotech/lilly-debuts-nvidia-supercomputer-fanfare-and-focus-escaping-traditional-pharma-lifecycle> (both cited Oct 13 2026)
173. Feldman A., *Roche Bought Thousands of Nvidia AI Chips to Speed Up Drug Development*. Forbes. Mar 16 2026. Available from <https://www.forbes.com/sites/amyfeldman/2026/03/16/roche-bought-thousands-of-nvidia-ai-chips-to-speed-up-drug-development/> (cited Aug 15 2026)
174. MIT NANDA initiative, *The GenAI Divide: State of AI in Business 2025* (lead author Aditya Challapally). Available from <https://storage.ghost.io/c/44/95/449506ca-034e-480f-9725-fcde08ef1cc1/content/files/2025/09/The-GenAI-Divide-STATE-OF-AI-IN-BUSINESS-2025.pdf> (cited Aug 15 2026)
175. Ryseff J. et al., *The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed*. Aug 13 2024. Available from [https://www.rand.org/pubs/research\\_reports/RRA2680-1.html](https://www.rand.org/pubs/research_reports/RRA2680-1.html) (cited Aug 2 2026)
176. Cuatrecasas, *Digital Omnibus on AI: How Does It Impact Employment Relations?* July 25 2026. Available from <https://www.cuatrecasas.com/en/global/labor-and-employment/art/digital->

[omnibus-on-ai-how-does-it-impact-employment-relations](#); Van Eecke P., Regenhardt B., *Digital AI Omnibus Delays Key Deadlines, Introduces New Rules*. Cooley. Aug 3 2026. Available from <https://cdp.cooley.com/digital-ai-omnibus-delays-key-deadlines-introduces-new-rules/>; Interface, *An Operational Framework for AI Literacy in the Workplace*. Policy Brief. Aug 25 2026. Available from <https://www.interface-eu.org/index.php/publications/ai-literacy-workplace-framework> (all cited Oct 13 2026). The Digital Omnibus on AI entered into force on July 27 2026.

177. BCG, Bedard J. & Beauchene V., *AI Transformation Is a Workforce Transformation*. Feb 4, 2026. Available from <https://www.bcg.com/publications/2026/ai-transformation-is-a-workforce-transformation> (cited Aug 2 2026)
178. Gartner, *Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027*, Jun 25 2025. Available from <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027> (cited Aug 2 2026)

## Chapter 3: AI Literacy Today

Chapter 3 of the AI Literacy Study asks one question about four groups across the globe that will decide whether AI makes economies and societies stronger or more fragile: *what do they actually know and understand about AI?* Not what they feel about it, and not how often they use it, but which concepts they recognize, which skills they rate in themselves, what they fear, what they want to learn, in what format, and what stops them.

edutAlnment designed and ran four original surveys between February and May 2026. The survey of *young people and students* (n=833 with 211 responses including open questions) covers the generation from sixteen through the first years of working life, mostly in the United States, Singapore and the UK. The survey of *teachers* (n=1,149 with 131 full responses) reached mid- and late-career educators, most of them US-based. The survey of *experienced professionals aged 45 and older* (n=1,065 with 221 full responses) is the only one where Europe dominates: three-quarters of respondents are based there. The survey of *small and mid-sized enterprises (SMEs)* (276 full responses from 915 surveys) draws mainly on the United States, UK, Singapore and Saudi Arabia; the European Union accounts for 5.1% of respondents. *In the coming months all surveys will be rolled out to further geographies - including continental Europe - and selected industries in collaboration with data collectors.*

Throughout, edutAlnment data are compared with secondary sources: surveys and reports by international organizations (OECD, UNESCO, IMF, the World Economic Forum), governments and national programs (Estonia, Singapore, the UAE, Beijing, Chile, South Korea, the United Kingdom), consultancies and industry (PwC, Deloitte, McKinsey, Microsoft, IBM, Verizon, Gallup), and researchers (RAND, Brookings, Stanford, MIT, Carnegie, Vivienne Ming, Barbara Oakley and others). Two reference points deserve mention because they set the standard against which any AI literacy work in 2026 must be read. The EU–OECD *AI Lit Framework*, finalized in June 2026, organizes AI literacy into four domains, Engage, Create, Manage and Shape, and will anchor the PISA 2029 assessment. Article 4 of the EU AI Act obliges every organization deploying AI in a professional context to take measures on the AI literacy of its staff, with enforcement from August 2026.

Useful as they are to ministries and standards bodies, existing AI literacy frameworks are not something a sixteen-year-old, a language or chemistry teacher, a fifty-year-old account manager or the owner of an accounting firm can pick up and learn from. They define competencies; they do not clarify what these involve and constrain, and they say almost nothing about what AI can and cannot do today, what it costs, and where it fails.

This study is different. Chapter 2 provides clarity on the current vulnerabilities of AI and the risks that follow from them, and on the cost drivers shaping AI implementations today, which any decision on AI strategy must weigh. This chapter shows where four populations actually stand. Both rest on more than survey data: on

200 interviews Anastassia Lauterbach recorded between January 2024 and September 2026, and on real AI implementation cases.

Judged against that body of work, edutAlnment occupies a specific position. The frameworks are top-down: they define what literacy *should* be. Our surveys in Chapter 3 are bottom-up: they ask what people *do* know. The two are complementary.

The limitations of edutAlnment surveys are equally clear. These are samples of engaged respondents, not probability samples, and they over-represent the United States, Singapore and the Gulf – those markets investing most in AI industries. Where our numbers can be checked against representative studies, RAND's youth panel, the Walton/GSV/Gallup survey, the OECD's Digital Education Outlook and D4SME survey, Microsoft's Work Trend Index, they land in the same range, which gives us confidence in the direction of the findings if not in their decimal points. *The near-absence of European respondents in three of four surveys is itself a finding*, and one we return to in the SME section.

Six themes run through every single population. They appear under different names in each section, the practitioner paradox for teachers, the deployment paradox for SMEs, but they describe one phenomenon:

- *Usage is near-universal; understanding is not.* Between 86% and 97% of respondents in each cohort use AI tools. Yet only 25.5% of students recognize the concept of AI bias, 11% of teachers describe their knowledge as advanced, 7% know what an open-weight model is, and the most requested training topic among SMEs, chosen by 59,1%, is *what is AI?* People know the product, not the mechanism, and a foundation that thin does not transfer when the product changes.
- *AI amplifies what is already there.* The path from education into work was failing before LLMs; cheating was endemic before ChatGPT; most cybersecurity breaches still trace to unpatched systems and mismanaged credentials. AI makes existing patterns and trends faster, cheaper and harder to reverse, which is why waiting to see is not an option.
- *Cognitive debt is the risk every cohort names in its own words.* Losing critical thinking is the top concern of students (44,7%), ahead of job loss. Teachers see it in their classrooms, SMEs see it as automation bias in junior staff, and professionals over 45 worry that tool fluency will erode the judgment they spent decades building. *AI amplifies skill only in those who already have it:* the goal, in Vivienne Ming's terms, is to build cyborgs who spar with the machine, not automators who copy from it or validators who seek AI information to confirm what they already believe.
- *The barriers are cost and time, not attitude.* Cost is the top obstacle for students (44%), teachers (62%) and the 45+ cohort (71%); time comes second. Only 7% of professionals over 45 cite lack of confidence with the technology. For SMEs the binding cost is not the license but the expertise and time needed to make a tool work inside a real business. These populations are not resistant from the behavioural perspective; they are structurally blocked.

- *The constraint sits in institutions, not individuals.* Only 34.5% of teachers feel their school supports them; 44% of the 45+ cohort receive no employer upskilling; students describe being “on our own”; 64% of SMEs tolerate shadow AI by default rather than by decision. Microsoft's finding that organizational factors carry twice the weight of individual mindset holds across all four groups.
- *The demand exists, and its shape is known.* Three-quarters of students, two-thirds of teachers and a cohort of professionals paying for AI subscriptions out of pocket are asking to be educated. They converge on the same design: fundamentals before tools, short modular formats, hands-on practice with feedback, and an emphasis on judgment, the skill employers now demand at the door and the one no benchmark measures.

**How to read this chapter.** The four sections that follow can be read independently. Each opens with the themes of its population, presents the edutAlnment data, tests them against secondary sources, and closes with recommendations. Each carries its own list of sources and secondary literature. Section 1 covers young people and students; Section 2 - teachers; Section 3 - experienced professionals aged 45 and older; Section 4 - SMEs.

## Section 1: AI Literacy in Young People and Students

The generation between sixteen and their first job did not "adopt AI" - it was born into it. We identified several themes describing the state of AI Literacy in young people:

First, *usage is near-universal; literacy is not*. In the edutAlnment survey of young people (n=833),<sup>1</sup> 94.2% have used AI tools and 45.0% use them daily or several times a week - yet awareness of AI bias and fairness reaches only 25.2%, and protecting privacy is the AI skill in which respondents feel least confident.

Second, *young people are asking to be taught, not protected*.<sup>1</sup> Across 219 open-text responses of the survey, the most consistent signal has been a request for institutional guidance: *AI literacy as a core competency, embedded across subjects, taught before tools are used*.

Third, *the entry-level labor market is being reshaped faster than the education system can respond*. Demand for AI skills in entry-level roles more than doubled in the past six months, and 35% of entry-level jobs now require them; yet employers simultaneously rank AI skills last among career-readiness competencies, behind communication, teamwork, and critical thinking.<sup>2</sup> Entry-level roles are undergoing what PwC calls *senior-isation - junior jobs now demand judgment, management, and initiative from day one* in employment.<sup>3</sup>

Fourth, *the cognitive debt issue is real*. MIT Media Lab research on cognitive debt, Brookings' yearlong 50-country study, and edutAlnment research on human-machine collaboration come to the same conclusion: the question is not whether young people use AI, but whether the relationship sharpens the mind or dulls it. Beneficial and negative impact on students' cognition can be observed.<sup>4, 5, 6</sup>

Fifth, it is *absolutely possible to introduce AI into education in a way that benefits students*,<sup>7</sup> enhances their existing abilities across different disciplines in science and the arts, and shows them new ways of working with data, research, productivity, and information organization. *Forbidding AI vs. blindly signing for LLM licenses on campus are two extremes, both damaging to students' chances to thrive in their future professional lives*.<sup>8</sup> The unfortunate truth is that the adults designing the curricula, hiring policies, and regulations have often a limited understanding of AI technologies.<sup>9, 10</sup>

Sixth, *the risks extend beyond the classroom or workplace*. A 2025 survey found 72% of US teens have used an AI companion; researchers documenting chatbot harms now argue that unsupervised use by minors is a safety issue, and a growing clinical literature shows that systems built to please can magnify psychiatric vulnerability.<sup>11, 12</sup>

Seventh, *the gender gap accelerates*. Young women report lower AI self-efficacy at comparable competence, are approximately 22% less likely to use generative AI (genAI) tools, and enter a labor market in which women are already 16 percentage

points less likely to have used genAI for work.<sup>13, 14</sup> The gap opens early: in the UK, male undergraduates are 12 percentage points more likely than female undergraduates to have used AI at school.<sup>15</sup>

Eight, *this is a global problem with unequal geographical distribution*. An estimated 1.2 billion young people in emerging and developing economies will reach working age over the next decade, entering economies where AI investment is booming but AI literacy infrastructure barely exists.<sup>16, 17</sup>

Nine, AI Literacy Programs are imperative to prepare young generation for the future they can thrive in.<sup>3, 17</sup> As AI becomes embedded in hiring pipelines, academic assessment, and workplace workflows, *the gap between those with guided AI literacy and those without will widen*.<sup>5, 18, 19</sup> The challenge is educational, political, and organizational,<sup>20</sup> not necessarily technological.

In the following pages, we focus on how young people use AI and what their first employers might demand from them. We particularly focus on how AI can do both: damage the thinking abilities of young people, and enhance their skills if approached as a tool to enable additional research and hypothesis testing in problem solving tasks. We close with recommendations for improving AI Literacy in the young generation. The extended whitepaper *AI Literacy in Young People and Students* includes further details on use of AI in formation of pre-school children, cognitive debt, gender gap, and mental health risks associated with the growth of AI companions.<sup>1</sup>

## **The Generation Between School and First Employment**

This study focuses on the generation of young people from sixteen years old through the first years of working life. Despite being digital-natives, the young generation remains significantly unprepared to use AI with the critical judgment over its capabilities and vulnerabilities, the safety awareness, or the strategic intention that the labour market demands.

The edutAlnment survey on AI Literacy in Students and Young People<sup>1</sup> was answered by 833 young people, mostly from three geographies. Among respondents who reported their country, 50.5% live in the United States, 25.2% in Singapore and 24.3% in the United Kingdom. With the latest survey wave, the UK moved from a marginal presence - only ten European students had responded to the first wave, a data point in its own right - to almost a quarter of the sample; respondents from continental Europe remain largely absent. The sample spans school pupils to young professionals: 11.1% are under 18, 71.8% are aged 18 to 29, and 16.7% are 30 or older. Gender representation was 62.2% female and 37.8% male. The largest single group was working full-time (42.5%), with 20.6% current undergraduates and 6.0% postgraduate students. The most common fields of study were STEM (20.6%), Business or Economics (19.0%), and Health or Medicine (15.1%).

Our respondents know AI as a consumer product because it reached them through app stores, social feeds, and homework experiments, not through curricula. The

most widely recognized AI concepts were large language models (LLMs) such as ChatGPT, Claude, and Gemini (67.8%), AI image or video generation (63.2%), and AI-powered recommendations in social media and shopping (50.0%). Awareness dropped sharply for the concepts underpinning ethical and advanced use of AI: AI bias and fairness was recognized by only 25.2%, and agentic AI by 22.7%.

A combined 57.4% of respondents describe themselves as having good or advanced AI understanding. Asked about specific skills, they feel most confident about using AI without becoming over-reliant on it (55.2% very or extremely confident) and writing clear prompts (53.6%) - a confidence in their own independence that sits uneasily with the 71.9% who worry that young people may become too dependent on AI. Protecting privacy sits at the bottom (38.1%), just below recognizing bias in AI responses (43.9%) - the skill that underpins responsible AI use at a societal level. 31.0% of respondents are not confident or only slightly confident protecting their privacy when using AI - a gap made more concerning by the volume of personal data young people feed into AI tools for academic and professional tasks. Our survey respondents know what they do not know. Asked which skills they most want to improve, they rank fact-checking AI answers first (48.1%), followed by AI for studying (35.1%), privacy and data safety (34.3%), and writing with AI without losing one's authentic voice (33.9%).<sup>1</sup>

A combined 55.7% of respondents hold a positive view of AI, 25.1% remain uncertain, and only 19.1% are negative. Our respondents love learning faster (55.3%), brainstorming (49.8%), communicating better (45.5%), and saving time (44.3%). Yet 40.9% feel pressure to use AI so they do not fall behind, and 74.0% expect AI to significantly change education and early careers within the next five years.<sup>1</sup>

The *leading concern* of our respondents is not job loss, not deepfakes, not even privacy. It is *losing critical thinking or creativity* - cited by 44.7%, ahead of false or misleading answers (41.3%), cheating and plagiarism (40.4%), and privacy (39.1%). AI replacing jobs ranks 28.5%. 71.9% agree that young people may become too dependent on AI, and 62.1% that AI makes it harder to know which information is trustworthy.<sup>1</sup>

This is a remarkable finding from our dataset. The dominant anxiety of the most AI-immersed generation in history is metacognitive: they are worried about what the tool is doing to their own minds. They have, in effect, independently arrived at the central concern of the cognitive-science literature.

Our survey sits within a fairly consistent 2025-2026 pattern of other studies: young people self-report roughly 40-70% concern that AI is weakening their thinking, with the worry peaking around the transition into work - while genAI usage keeps rising.

The RAND American Youth Panel is probably the closest match to edutainment findings.<sup>21</sup> In 2025, most U.S. students who used AI for homework worried about its effects on critical thinking, based on 1,214 youth aged 12 to 29 surveyed in December 2025, and nearly 7 in 10 middle and high school students said they were

concerned that using AI for schoolwork is eroding their critical thinking skills. But young people kept using it anyway - high schoolers' homework use rose from 49% to 60% between May and December 2025.<sup>21</sup> The Walton Family Foundation / GSV Ventures / Gallup comes to the similar conclusions, with 49% of 13- to 28-year respondents from 2025 and respectively 42% in 2026 believing AI might harm their critical thinking skills.<sup>22</sup>

Newer data reinforce the pattern. In February 2026, Pew Research Center found 64% of US teens using AI chatbots and 54% using them for schoolwork; among teens expecting AI to have a negative effect on society, overreliance and the loss of critical thinking or creativity was the most-cited reason (34%), ahead of job loss (25%).<sup>23</sup> Common Sense Media reported in August 2026 that 70% of US teens use AI for schoolwork and nearly two-thirds of them say they offload critical thinking to it - while only 27% say a teacher has explained what AI is and how it works.<sup>24</sup> Among US adults aged 18 to 29, 55% are now more concerned than excited about AI in daily life - up from 47% in 2025 and the first majority on record - and only 11% are more excited than concerned.<sup>25</sup> Self-assessment differs markedly between samples: only 31% of US 18- to 29-year-olds told Pew they are very or extremely confident using AI chatbots,<sup>26</sup> while 57.4% of our respondents claim good or advanced AI understanding - an expected effect of a self-selected sample interested enough in AI to open a survey about it. The core anxiety, however, is identical.

In their own learning, our respondents are mostly careful users: 45.1% use AI for support but do the final work themselves, 25.8% use it only for brainstorming or explanations, 12.0% sometimes rely on it heavily, and 13.3% avoid it completely. 52.8% have seen AI affect learning in their school or college both positively and negatively. The guidance they want most is how to fact-check AI responses (55.8%), how to use AI without plagiarizing (52.8%), and how to cite or disclose AI use (44.2%); 60.0% would like clearer guidance on ethical AI use, 69.4% want AI tools to be more transparent about how they generate answers, and 58.7% say schools, universities and employers should teach AI literacy as a core skill.<sup>1</sup> UK undergraduates surveyed by HEPI tell a similar story: 95% use AI in some form and 68% consider AI skills essential, but only 48% feel that teaching staff help them develop those skills.<sup>15</sup>

In the edutAlnment survey, 74.2% of respondents express at least moderate interest in a structured AI literacy program, with 43.1% very or extremely interested. Their preferred topics blend the practical and the protective: AI for studying (50.2%), fact-checking and verifying AI output (49.8%), creativity (47.1%), writing and communication (46.7%), and prompting (44.4%). Preferred formats point to hybrid, modular design: live online workshops (36.0%), in-person workshops (30.7%), video lessons (29.8%), short self-paced modules (29.3%), and peer learning (28.4%). The single greatest barrier is cost (44.0%), followed by lack of time (38.2%) and uncertainty about usefulness (32.9%). What would make a program most valuable? Time-saving and quality tips (50.7%), career-focused outcomes (43.6%), real examples relevant to their lives (43.1%), and hands-on practice with feedback (42.2%). Certificates matter too (34.2%): this generation understands credentials as currency.<sup>1</sup>

Asked how schools and universities should integrate AI, respondents agreed on five themes: *foundational AI literacy before tool use; ethical integration across all subjects rather than quarantine inside computer science; candid acknowledgment that teachers are unprepared* (“We are on our own,” as one respondent put it); *hands-on, project-based learning over passive instruction; and responsible use instead of prohibition* -because, as another respondent observed, “warning does not translate into understanding.”<sup>1</sup>

The mandate for policymakers and educators could hardly be clearer. *The customer for AI literacy exists, knows what it needs, and is waiting for institutions to show up.*

## **Europe and Other Regions: What Is Different?**

For young people, the difference between Europe - including the UK - and other regions lies less in adoption than in attitudes and, above all, in institutional support.

Our data allow an indicative comparison: the roughly 50 respondents added in the latest survey wave are predominantly UK-based (UK respondents rose from a handful to 51), and the change in aggregate results shows how they differ from the earlier, mostly US and Singaporean cohort.<sup>1</sup> The UK group was somewhat less likely to have used AI at all (about 88% against 96%), yet those who use it do so more intensively (about half daily or several times a week, against 43%); it is more positive about AI (about 61% against 54%), noticeably less aware of AI bias (about 17% against 27%), less worried about losing critical thinking (about 35% against 47%) and more worried about false or misleading answers (about half against 39%) - and slightly less eager to join an AI literacy program (68% against 76% at least moderately interested). Third-party data confirm that usage among young Europeans is as high as anywhere: 63.8% of EU 16- to 24-year-olds used generative AI in 2025, ranging from 44.1% in Romania to 83.5% in Greece,<sup>27</sup> and 79% of UK 16- to 24-year-olds use AI tools such as ChatGPT, Copilot or Gemini, against 54% of all UK adults.<sup>28</sup> Scepticism is rising fastest in the US, where the share of under-35s more concerned than excited about AI climbed from 40% to 55% in two years - but Pew's 36-country study shows young Europeans moving in the same direction, with Swedish under-35s jumping from 24% to 45% in a single year.<sup>29</sup>

The real divergence is in what institutions do. Singapore has integrated AI literacy into its curriculum, co-curriculum and self-directed learning resources, updated its Cyber Wellness lessons to cover validating information produced by generative AI and identifying deepfakes, and will make an AI-strengthened Code for Fun programme available to all schools in 2027;<sup>30</sup> in Singapore, more than half of eight-year-olds have already used AI.<sup>31</sup> England announced in November 2025 that AI literacy will become part of its reformed school curriculum;<sup>32</sup> in the EU, AI literacy is a legal obligation for organizations deploying AI - recast in July 2026 as a duty to support its development - but the obligation does not reach classrooms;<sup>33</sup> and in the US, only 27% of teens say a teacher has explained to them what AI is and how it works.<sup>24</sup> Across all regions, the gap is the one our respondents describe: high use, rising doubt, and too little instruction.

## Professional Thriving and AI Literacy

It is tempting to blame every disruption of the labour market on AI, but the system for moving students from schools and universities into careers was failing well before November 2022, when ChatGPT was released to the general public. The same finding goes across all areas of economy and society. AI (and particularly genAI) magnifies and speeds up existing trends and patterns - positive and negative. This is why the "wait and see" approach regarding democratizing AI Literacy across the whole population is not an option.

Research on career regret across more than 10,000 professionals at various career stages found that nearly six in ten said they would do things differently if they could start over, and more than 40 percent would choose an entirely different occupation.<sup>34</sup> Gallup's global workplace data echoes the problem: 64% of employees globally are not engaged (the group Gallup associates with quit quitting), with another 16% actively disengaged.<sup>35</sup> These findings challenge how the old conveyor belt worked out for those on it, and they set the stakes for the young people now stepping onto it. The conveyor belt runs along statements like these: "Pick the right schools. Choose the right subjects. Stack the résumé. Land a job within a proven set of safe fields."

## University Education in the Age of AI

One broken link of the conveyor belt is, of course, university education. University, while delaying the transition into working life, helps build independence and personality. But this is linked to high financial burden. There is a need to recalibrate - not to abandon higher education, but to stop treating it as a guarantee of employability and start asking what, specifically, it teaches that a changing labour market will reward.

*The university sector faces a clear task: re-earning the confidence it is losing. For example, when a third of the English public calls university a waste of time while enrollment hits historic highs,<sup>36</sup> the institution is relying on default choices, not demonstrated value.<sup>37</sup> Embedding rigorous, career-connected AI literacy curriculum and AI practice into every discipline is one step to demonstrable value.*

Another is the *convergence of AI literacy and entrepreneurship education*.<sup>38, 39</sup> Generative AI tools have collapsed the distance between an idea and a working product: AI app builders now let a student describe a problem in plain language and ship a functioning application within a week, with no coding background required. That collapse turns students from consumers of technology into creators of it - and reframes AI, in the words of one entrepreneurship-education framework, from a threat to be feared into a co-founder, a competitor, and a tool.<sup>40</sup>

Real examples are no longer rare: high-schoolers are launching AI-powered study tools, and graduate-student teams now compete for six-figure prizes at university AI-startup challenges built around genuine customer discovery. Andrea Olsen, a Caltech undergraduate who contributed to the design of edutAlnment surveys, launched an AI company at the age of 19. Another contributor to our AI Literacy

Study, Elisa Torres Durney, a Duke undergraduate, created her non-profit *Girls in Quantum* at the age of 16. The institutional infrastructure has moved quickly to support founders like Andrea and Elisa - a majority of US AI bachelor's programs now integrate entrepreneurship coursework, and dozens of universities pair AI degrees with startup-incubator access.<sup>38, 39, 40</sup>

The honest question is whether building AI businesses is actually a path to employment and thriving, or merely an appealing story. A fifteen-year study of more than 64,000 university graduates found that those who tried entrepreneurship soon after graduation, before accumulating much labor-market experience, enjoyed a wage premium relative to peers who never founded anything, and suffered no significant penalty on returning to traditional employment; the wage penalty falls only on those who interrupt an established career to start a business late.<sup>41</sup> A field experiment on hiring reinforces the point: a successful founding experience sends a strong quality signal to employers, validating capabilities that early-career candidates otherwise struggle to demonstrate.<sup>42</sup> For young people locked out by the seniorisation of entry-level roles, creating a business may be one of the few remaining ways to accrue the judgment and initiative that employers now demand at the door.

A realistic approach by governments and educators would treat entrepreneurial capabilities as an asset rather than as a remedy for youth unemployment. The global venture failure rate is roughly 70%, higher still for student ventures - and only a minority of student founders try again.<sup>43</sup> Still, *entrepreneurial skills - problem-finding, customer discovery, building and shipping with AI, comfort with failure - are durable* and represent something the traditional conveyor belt never planned for.

## Entering the Labour Market in the Age of AI

AI skills are simultaneously the fastest-growing requirement and the least-valued competency. The resolution of this *paradox is not that AI skills do not matter; it is that employers treat them as necessary but insufficient - a hygiene factor layered on top of the deeply human capabilities that actually differentiate candidates*. Employers seeking AI skills, ask for identifying the right tool for the task (75.0%), developing effective prompts (72.7%), and - critically - analyzing and revising AI outputs (65.9%). That third item is precisely the fact-checking competency young people themselves rank as their top training priority.<sup>2</sup>

PwC's 2026 Global AI Jobs Barometer, describes *seniorisation*: roles requiring zero to two years of experience now routinely demand capabilities traditionally associated with mid- or late-career stages - management, strategy, decision-making. The new requirements concentrate in interpersonal skills and deeply human attributes: creativity, judgment, initiative, teamwork, along with increased demand for face-to-face presence. Employers, notably, add these skills to requirement lists rather than swapping new ones for old.<sup>3</sup>

Junior roles once existed precisely so that judgment could be formed on the job; now judgment is demanded at the door. Greater demand for initiative and management skills even in junior roles suggests AI-adopting firms are seeking workers who can be

trusted to make good decisions on the fly, and not only to initiate but to progress and manage pieces of work from day one. Most exams and assignments, meanwhile, do relatively little to assess or reward creativity, agency, and good judgment.

UK data from HESA and Prospects, surveying almost 200,000 graduates fifteen months after completing 2022–23 degrees, found computer science graduates more likely to be unemployed than those from any other subject: nearly one in ten, comfortably above the 6.2% graduate average - and above performing arts (5.2%), English literature (6.4%), and marketing (7.2%).<sup>44</sup> Entry-level software-engineering postings have fallen between 30 and 40% since early 2023, depending on the geography. There are now approximately four times as many advertisements for senior tech roles as junior ones in the UK.<sup>45</sup> Close to a quarter of a million tech employees worldwide lost their jobs in 2025 alone, even as Microsoft's chief executive disclosed that 30% of the company's code was already being written by AI.<sup>46, 47</sup>

Behind the statistics stand young graduates like the first-class graduate who applied to roughly one hundred junior developer roles, progressed twice, and received no feedback at all - before finally being hired thanks to a family friend's endorsement post.<sup>48</sup>

The recruitment process itself has become an AI battleground. One in eight UK job hunters had already made fifty applications by February of their final university year; law applications now routinely involve a minimum of five stages, with candidates immediately tested and never speaking to a human.<sup>48</sup> Automated recruitment lets candidates fire off hundreds of applications without hearing back from a single person.<sup>49</sup>

Employers see the mirror image: 44% report detecting AI-generated applications, while only 20% use tools to detect AI assistance in applications, testing, or interviews.<sup>2</sup> Both sides are now uncertain which humans they are actually evaluating.

Meanwhile, the *most sophisticated employers are redesigning assessment around the new reality rather than against it*. McKinsey - which has deployed some 20,000 AI agents alongside its roughly 40,000 staff - is piloting graduate recruitment in which candidates use its AI assistant, Lilli, to complete tests designed to reflect consultants' new ways of working.<sup>50</sup> The test is no longer what you know; it is how you work with the machine while solving McKinsey-style case studies - their traditional hiring practice. That is a definition of expertise the education system has barely begun to teach.

Young people see these shifts clearly. Asked about AI in hiring and early careers, our respondents worry most that AI makes their applications feel generic (57.2%), that AI screening may be unfair or unclear (55.9%), and that AI may reduce entry-level opportunities (51.5%); 35.8% fear employers expect AI skills they have not yet learned, and 30.6% do not know how to talk about their AI use professionally. Only 37.6% are very or extremely confident using AI in professional contexts.<sup>1</sup> They use it

nonetheless - for workplace writing (36.7%), CV writing and interview preparation (36.2% each), and cover letters (31.9%) - and what they hope AI will do above all is help them learn skills faster (53.3%) and be more productive at work (50.2%).<sup>1</sup> Their worry about entry-level opportunities is well founded. Stanford's August 2026 update of its "Canaries in the Coal Mine" analysis finds employment of 22- to 25-year-olds in the most AI-exposed occupations about 19% below its counterfactual by June 2026, up from 15% a year earlier, with the adjustment coming mainly through reduced hiring;<sup>51</sup> UK graduate job postings in July 2026 were about 7% lower than a year earlier and at their lowest level since 2020, even as a record 9.4% of all UK postings mentioned AI;<sup>52</sup> global youth unemployment rose to 12.4% in 2025, although the ILO considers it too early to attribute the decline of entry-level office jobs to AI;<sup>53</sup> and 73% of US adults aged 18 to 29 now expect AI to lead to fewer jobs over the next twenty years.<sup>25</sup>

### **Cognitive Debt or Dividend?**

The question beneath every AI literacy debate is not whether young people will use AI - they already do - but what using AI does to the mind on the other side of the screen. We have already touched on the problem of cognitive debt in Chapter 2, in our discussion of AI risks. In edutainment interviews, we see that the mindless introduction of genAI causes harm, while designing personalized educational tools enhances students' performance and capabilities. These findings are supported by researchers and educators with over a decade of experience using technology for problem-solving and research.

Drawing on 505 interviews with students, teachers, parents, and technologists, hundreds of studies, and a Delphi panel, the Brookings Institution's report identifies six risk clusters - to cognitive development, social and emotional development, trust in education, safety, autonomy and agency, and equity - and concludes that, at this point in AI's trajectory, the risks of utilizing AI in education overshadow its benefits, largely because the risks undermine children's foundational development while the benefits merely enhance its surface. The same report states that well-designed AI anchored in sound pedagogy can optimize teacher time-spent, personalize learning, and extend education to neurodivergent students and students with disabilities.<sup>5</sup> Lack of AI Literacy in teachers is one of the obstacles to deliver positive AI outcomes in learning despite some positive examples when AI is used as a personalized learning tool.<sup>9</sup>

Why can AI do both, produce negative cognitive outcomes and greatly support problem solving tasks?

The computational neuroscientist Vivienne Ming has documented this divergence for the better part of two decades, and the same pattern shows up everywhere: two people use the same tool for the same task; one brain gets sharper, the other goes quiet, its neural state drifting toward something closer to watching television than solving problems.

Ming sorts technology users into three types: automators, who copy and paste; validators, who seek confirmation of what they already believe; and cyborgs, who spar with the machine. The differentiator is the willingness to choose productive discomfort - being wrong, or doing the unglamorous work of interrogating an answer that looks correct.<sup>6</sup>

Ming had participants - smart students with no domain expertise — predict genuinely difficult outcomes, such as oil prices six months out, and measured them against professional forecasters with real money at stake. The overwhelming majority simply handed the AI their question or asked it to confirm a conclusion they had already reached; those groups performed no better than the AI alone, or worse, and gamma-band brain activity showed them working, in her words, like people watching television rather than thinking. But a minority (Ming calls them *cyborgs*), while provided with even a small open-source model, produced predictions better than the best humans, better than the best AI, and as good as the professional forecasters.<sup>6</sup> What predicted who became a cyborg was not any AI benchmark but fluid intelligence, curiosity, intellectual humility, and perspective-taking - whether, when the AI handed over a satisfying answer, the person kept looking, and whether, when it said they were wrong, they engaged rather than ignored it.

Between 2012 and 2016, Norma Ming and Vivienne Ming analyzed the discussion-forum activity of roughly 60,000 students across undergraduate and MBA programs. The students who earned top marks were those who were wrong most often in the forums: they proposed, explored, and were frequently and visibly incorrect. The nine out of ten students who passed safely were far less likely to venture a claim they could not defend.<sup>6</sup> Error, publicly risked and corrected, is how expertise forms - and, as Ming notes, one genuine benefit of AI is that it can tell a student they are wrong without social pressure, offering a platform to say “here is what I believe; tell me why I’m wrong.” An always-available machine that supplies defensible answers on demand removes that formative friction unless educators deliberately design it back in.

This is also what AI benchmarks miss when frontier labs announce near-perfect scores on coding, law, or medical exams. Those exams test models in isolation. In the field, AI works alongside a person — and the relevant performance is exactly this combination, not the model on its own. The point generalizes beyond the classroom: as Ming warns, a risk-averse manager taking answers from a super-intelligent machine should remember that the same answers are in every competitor's hands, so an organization that does not differentiate on something other than AI will be disrupted.<sup>6</sup>

## **AI Bans in Education**

Avoiding the AI as a topic or forbidding it in schools or universities is not helpful, nor does it prepare the young generation for the future.

AI bans fail even on its own terms: enforcement is unreliable because AI-detection tools are imperfect and students can use rewriting tools to make AI-generated text

harder to identify, which pushes teachers into policing rather than teaching. False accusations arising from detector “false positives” poison relationships and put students on the defensive even when they have done nothing wrong. The Brookings Global Task Force on AI in Education documents the deeper cost: AI threatens to fray the relational bonds essential to education - many teachers distrust the authenticity of student work while students question whether their teachers' feedback is genuinely their own - and this erosion of trust can provoke cynicism that undermines the relationships on which meaningful learning depends.<sup>5</sup> The student voices in the edutAlnment survey describe the institutional response to date as abdication - bans, warnings, and detection theater in place of instruction. One European respondent put it starkly: *“We want to learn, but our teachers behave as if AI wasn't there. So we are forced to learn from TikTok and Instagram.”*<sup>1</sup> An education system that responds with silence and suspicion does not keep children safe from companions, deepfakes, algorithmic feeds, and AI-mediated hiring; it merely teaches them that AI is a vice to be hidden rather than a power to be governed.

Educating students about what AI is and how it works must therefore start as early as middle school - a stage researchers identify as ideal because of students' developmental readiness for abstract thinking. This does not mean *tool training* - which “buttons” to push on a given user interface, as such training rushes rollouts serving only AI vendors. Any serious AI Literacy course should explain how AIs differ from humans and why that difference matters, provide an overview how AI models are being designed and trained, and what are differences between foundational and open-weight LLMs.

The experience of edutAlnment and the “Romy&Roby” AI series shows even 12 years old are hungry for learning about AI concepts: one class was captivated by Dr Lauterbach's comparison of behavioural patterns in the Nematode *C. elegans* (one of the most studied worms) with the Roomba, one of the first commercial robots; sixth-graders in Maastricht asked how the granular prefrontal cortex enabled Theory of Mind - a capability that is still absent from any LLM - and why Google's Gemini model produced historically inaccurate images of the U.S. Founding Fathers. After three one hour sessions, young, environmentally conscious students were able to explain why current LLMs are “that power-hungry.” Students want to learn and understand, but they need great stories, plain language, and time to ask questions.

## **Recommendations: AI Literacy in Young People and Students**

This report has been critical of institutions. But a report that ended there would repeat the very mistake it documents: treating young people as passive recipients of whatever adults build or fail to suggest. Young people have agency, and some of them are already using it in ways worth studying and imitating. This section sets out four recommendations for young people, followed by an honest account of what adults owe them in return.

## **1. Take Your Own Future Seriously**

The first recommendation is not a program but a call: take your own future seriously. Do not outsource it to institutions that are visibly struggling to keep pace. The survey data show that young people already know what they lack - fact-checking ranks first among the skills they want to improve (48.1%), followed by using AI for studying and by privacy and data safety.<sup>1</sup> Taking your future seriously means acting on that self-knowledge now, without waiting for a curriculum committee, a ministry, or an employer to hand you a syllabus. It *means treating AI literacy the way earlier generations treated a driving license*: a basic competence for adult life that you go and get, deliberately, before you need it. It also means resisting the two easy postures the current moment invites - cynical detachment ("it's all hype") and passive consumption ("the tools will do it for me.")

## **2. Learn How AI Works, Not Only How to Use It**

The durable AI Fluency skills are require understanding how generative systems produce plausible falsehoods, recognizing bias, protecting privacy, knowing when not to use genAI at all. None of this requires a computer science degree.

Here are some selected resources:

- *Elements of AI* (University of Helsinki and MinnaLearn) - a free online introduction to AI for non-experts, available in many languages and completed by learners in over 170 countries.<sup>54</sup>
- *Day of AI* (MIT RAISE) - free, hands-on AI literacy curriculum for K-12 students, with ready-to-teach units on how AI works, data science, and responsible use, plus teacher training.<sup>55</sup>
- *AI4ALL Open Learning* - a free, self-paced, project-based program introducing high school students to AI and its ethical implications, open to students worldwide.<sup>56</sup>
- *Machine Learning for Kids* and *Experience AI* (Raspberry Pi Foundation with Google DeepMind) - hands-on lessons in which learners build and train their own simple models, designed for roughly ages 8-14.<sup>57</sup>
- *Romy&Roby* - Anastassia Lauterbach's book series, beginning with *Romy, Roby and the Secrets of Sleep*, which explains AI, robotics, and quantum technologies to children and their parents through storytelling, with a "Science Behind" appendix for curious minds. *AI Snacks with Romy&Roby* - the accompanying podcast, which translates AI and robotics from complex science into accessible discussions for teens, parents, and teachers, alongside the edutAlnment and Romy&Roby platforms.<sup>58</sup>

The list is deliberately mixed: university courses next to a children's book series. AI literacy is a family and community project, and the best learners often teach what they learn to someone younger.

### **3. Ask Adults and Companies for Knowledge - and Be Persistent**

A striking pattern connects the young people featured in this report: they asked. They wrote to companies, researchers, and executives - not for jobs or favors, but for knowledge - and enough adults said yes to change their trajectories.

Andrea Olsen, back then a US high-schooler interested in medical research, reached out directly to Alex Zhavoronkov, the CEO of the AI drug-discovery company Insilico Medicine. Zhavoronkov connected her with another young researcher, and the collaboration led to a peer-reviewed publication identifying new therapeutic targets for glioblastoma, the most aggressive malignant brain tumor.<sup>59</sup>

The sixth-graders from Maastricht described belong to the same pattern at classroom scale: given an adult willing to answer, students ask questions about worms and Roombas, hallucinating models and human brains.

Persistence and human agency matter regardless of age: Elisa Torres Durney, whose story follows below, recalls that in building her organization, her team was lucky if one expert in hundreds of emails replied - and that many adults simply did not take a teenager seriously.<sup>60</sup> Young people should ask anyway, repeatedly; the evidence says that a single yes can kick-start a career.

### **4. Build a Community and Teach What You Learn**

The strongest form of learning is teaching, and the strongest antidote to the institutional vacuum is community. When no institution offers the literacy you need, you can build the institution - as the following case study shows.

#### *Case study: Girls in Quantum*

Elisa Torres Durney grew up in Chile with a passion for science that ran from a childhood microscope to physics and engineering. In 2021, during the pandemic lockdowns, she won a full IBM scholarship to an eight-month Introduction to Quantum Computing program - her first formal course in the field, taken through IBM's *Qubit x Qubit*.<sup>61</sup> The course did two things at once: it revealed to her the potential of an emerging technology, and it showed her how few women and girls were in the room. When the program ended in April 2022, she refused to let her learning journey stop. She founded *Girls in Quantum*, a youth-led nonprofit providing free, multilingual educational resources on quantum computing to students worldwide, with a particular focus on girls from underrepresented communities.<sup>60</sup>

What began as videos and blog posts explaining qubits and quantum cryptography grew into a global platform: Girls in Quantum has reached more than 8,000 students

across 60 countries, built partnerships with institutions including IBM and UNESCO, and co-hosted the largest Spanish-language quantum hackathon in Latin America. Torres Durney - named a Top 10 finalist for the Global Student Prize in 2023 - now pursues an engineering degree at Duke University while continuing to lead the organization; she also contributed to the research underlying this report.

The lesson generalizes beyond quantum computing. A 16-year-old with one scholarship course and no budget built literacy infrastructure that reached sixty countries. Every reader of this AI Literacy Study can run a smaller version of the same play: start a club, translate what you learn into simple language (and your native language), teach a younger sibling, publish your notes on social media.

## **What Adults Owe Young People**

The young generation deserves leadership from those currently in charge - and, on the evidence assembled in this report, that leadership is failing them. It fails them when:

- schools respond to AI with bans, warnings, and detection theater instead of a real dialogue;
- employers demand day-one judgment while dismantling the junior roles in which judgment was traditionally formed;
- AI makers optimize companion products and ship them to minors without guardrails;
- the adults who write curricula and regulations do not understand the technology they are governing.

The failure is not about lack of resources. It is about lack of willingness to engage. *Every university campus and every school can foster greater knowledge and understanding of AI among youth* - without waiting for a ministry, a budget line, or a standardization institution. A faculty member can sponsor a student AI club. A school can run an AI literacy week using free online resources. A university can embed rigorous, career-connected AI practice into every discipline. The materials exist, the demand is documented, and the survey respondents are, in their own words, waiting for institutions to show up.<sup>1</sup>

Where institutions have not shown up, some parents have taken the matter into their own hands.<sup>62</sup>

That, in the end, is the fair bargain this report proposes. Young people: take your future seriously, learn the fundamentals, ask persistently, risk being wrong, build communities, celebrate each other's excellence. Adults: match that seriousness with instruction instead of prohibition, apprenticeship instead of gatekeeping, and products designed for development instead of engagement. Young people can do a great deal. They should not have to do it alone.

### Sources to Chapter 3\_Section 1

1. edutainment Survey: *AI Literacy, Young People and Students* (2026). Primary survey instrument, n=833, status October 2026; percentages refer to the respondents answering each question; 219 respondents answered the open questions. Analysis including confidence scales (Q12–Q14), attitudes (Q15–Q19), learning (Q20–Q23), career preparation (Q24–Q27), program preferences (Q28–Q32), and open-text responses (Q33–Q34). Survey design contributions by Andrea Olsen and Elisa Torres Durney. Extended whitepaper *AI Literacy in Young People and Students* available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Oct 12 2026). Original Survey available upon request.
2. National Association of Colleges and Employers (2026). *Job Outlook 2026, Spring Update. Survey of 185 employers*, Feb. 12 – Mar. 17, 2026. Available from <https://www.naceweb.org/research/reports/2026/job-outlook/spring-update/> (cited Aug 15 2026)
3. PwC (2026). *Global AI Jobs Barometer 2026. Analysis of more than one billion job advertisements across six continents*. Available from <https://www.pwc.com/gx/en/news-room/press-releases/2026/pwc-2026-ai-jobs-barometer.html> (cited Aug 15 2026)
4. Kosmyna, N. et al., *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*. MIT Media Lab, 2025. Available from <https://arxiv.org/abs/2506.08872> (cited Aug 15 2026)
5. Burns, M. et al., *A New Direction for Students in an AI World: Prosper, Prepare, Protect*. Brookings Institution, Center for Universal Education. January 2026. Available from <https://www.brookings.edu/articles/a-new-direction-for-students-in-an-ai-world-prosper-prepare-protect/> (cited Aug 15 2026)
6. Ming, V. , Episodes 76 and 77 of *AI Snacks with Romy&Roby* (“*Can Humans Compete With Machines with Dr. Vivienne Ming, the ‘Mad Scientist,’ Part 1 and 2*”).
7. Miller K. et al., *AI Tutoring Outperforms In-Class Active Learning*, Nature, 2025. Available from <https://www.nature.com/articles/s41598-025-97652-6> (; Wang, et al., *Tutor CoPilot* (2024). Annenberg Institute. Brown University. Available from [https://edworkingpapers.com/sites/default/files/ai24\\_1054\\_v2.pdf](https://edworkingpapers.com/sites/default/files/ai24_1054_v2.pdf) (cited Aug 15 2026)
8. Spector C., *What Do AI Chatbots Really Mean for Students and Cheating? Stanford Education Scholars Vicor Lee and Denise Pope Discuss Ongoing research Into Why and How Often Students Cheat*. Stanford Graduate School of Education. Available from <https://ed.stanford.edu/news/what-do-ai-chatbots-really-mean-students-and-cheating> (cited Aug 15 2026)
9. Lauterbach A., *AI Literacy in Teachers*. Whitepaper, October. 2026, Available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Oct. 13 2026).

10. *Survey: Gap Exists Between Students, Educators When it Comes to Using, Understanding AI*, July 22 2026. Available from <https://www.deseret.com/utah/2026/07/22/survey-gap-exists-between-students-educators-when-it-comes-to-using-ai/> ; Morin S. et al., *Teacher Optimism on AI is Dropping. The Reason Isn't the Technology*. Aug 6 2026. Available from <https://edtechinsiders.substack.com/p/teacher-optimism-on-ai-is-dropping> (both cited Aug 15 2026)
11. Haidt, J. et al. (2025). *Don't Give Your Child Any AI Companions*. After Babel, November 2025. Available from <https://www.afterbabel.com/p/dont-give-your-child-an-ai-companion> (cited Aug 15 2026)
12. Newport, C. (2026). *What's Worrying Jonathan Haidt Now?* Newsletter essay, Jan. 19, 2026. Available from <https://calnewport.com/whats-worrying-jonathan-haidt-now/> (cited Aug 15 2026)
13. Otis, N.G. et al., *Global Evidence on Gender Gaps and Generative AI*. Harvard Business School Working Paper. 2025. Available from [https://www.hbs.edu/ris/Publication%20Files/25-023\\_be8fb517-3dd5-40aa-97f9-4e42e1c8e6ff.pdf](https://www.hbs.edu/ris/Publication%20Files/25-023_be8fb517-3dd5-40aa-97f9-4e42e1c8e6ff.pdf) (cited Aug 15 2026)
14. Humlum, A. et al., *The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities Among Workers*. Proceedings of the National Academy of Sciences (PNAS) 122(1): e2414972121. Dec. 30 2024. Available from <https://www.pnas.org/doi/10.1073/pnas.2414972121>; Humlum, A. et al., *Large Language Models, Small Labor Market Effects*. Becker Friedman Institute Working Paper 2025-56, University of Chicago. Apr. 2025. Available from [https://bfi.uchicago.edu/wp-content/uploads/2025/04/BFI\\_WP\\_2025-56.pdf](https://bfi.uchicago.edu/wp-content/uploads/2025/04/BFI_WP_2025-56.pdf) (both cited Aug 15 2026)
15. HEPI and Kortext, *Student Generative AI Survey 2026*. HEPI Report 199. Mar 12 2026. Available from <https://www.hepi.ac.uk/wp-content/uploads/2026/03/HEPI-Report-199-Gen-AI-Survey-2026.pdf> (cited Oct 12 2026). Survey of 1,054 UK full-time undergraduates: 95% use AI in some form; 68% say AI skills are essential; only 48% feel teaching staff help them develop AI skills; men are 12 percentage points more likely than women to have used AI at school.
16. World Bank Group (2026). *Global Economic Prospects*, January 2026 (pre-Iranian war baseline: global growth of 2.6% projected for 2026). Available from <https://openknowledge.worldbank.org/entities/publication/bb904ec6-730f-4dd9-b1af-ad3153ee1616>; *World Population Prospects 2024* (database available from <https://population.un.org/wpp/>); Papagianni, V. et al., 2026. Available from <https://openknowledge.worldbank.org/server/api/core/bitstreams/a9e24256-baf8-45bb-9075-75e437e1d6f7/content> ; The Global Economy In Five Charts, World Bank Let's Talk Development Blog, Jan. 13, 2026. Available from <https://blogs.worldbank.org/en/developmenttalk/the-global-economy-in-five-charts3>; International Monetary Fund (2026), *World Economic Outlook, April 2026: Global Economy in the Shadow of War*. Available from

- <https://www.imf.org/en/publications/weo/issues/2026/04/14/world-economic-outlook-april-2026> (all cited Aug. 15 2026).
17. World Bank Group, Chrimes T. et al, The Global Jobs Challenge. Available from <https://openknowledge.worldbank.org/server/api/core/bitstreams/021e6eaf-4067-4494-b82a-fe1e86dd99fe/content>; Okolo C. et al, The Global Majority in International AI Governance, Available from <https://arxiv.org/pdf/2601.17191> (cited Aug 15 2026)
  18. Miao F. et al., *AI Competency Framework for Students*, 2024. Available from <https://www.unesco.org/en/articles/ai-competency-framework-students> (cited Aug 15 2026).
  19. Dusseault B. et al., *How State Policy Can Help Teachers Use AI Well*. State Education Standard. The Journal of the National Association of State Boards of Education. May 2026, Vol 26, No. 2. Available from <https://www.nasbe.org/how-state-policy-can-help-teachers-use-ai-well/>; OECD. *The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education*. OECD Artificial Intelligence Papers. Aug 2024, No. 23. Available from [https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/08/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education\\_0d7e9e00/15df715b-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/08/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_0d7e9e00/15df715b-en.pdf); Tolosa L. et al., *Integrating AI in Education: Navigating UNESCO Global Guidenlines, Emerging Trends, and Its Intersection with Sustainable Development Goals*. Research Gate. March 2025. Available from [https://www.researchgate.net/publication/389936132\\_Integrating\\_AI\\_in\\_Education\\_Navigating\\_UNESCO\\_Global\\_Guidelines\\_Emerging\\_Trends\\_and\\_Its\\_I](https://www.researchgate.net/publication/389936132_Integrating_AI_in_Education_Navigating_UNESCO_Global_Guidelines_Emerging_Trends_and_Its_Intersection_with_Sustainable_Development_Goals) [ntersection\\_with\\_Sustainable\\_Development\\_Goals](https://www.researchgate.net/publication/389936132_Integrating_AI_in_Education_Navigating_UNESCO_Global_Guidelines_Emerging_Trends_and_Its_Intersection_with_Sustainable_Development_Goals); Toppo G., *Four Takeaways from New Report on AI's Risks in Education*. Jan. 20 2026. Available from <https://www.the74million.org/article/four-takeaways-from-new-report-on-ais-risks-in-education/> (all cited Aug 15 2026)
  20. Li J. et al., *Overcoming the Organizational Barriers to AI Adoption*. HBI. Nov 11 2025. Available from <https://hbsp.harvard.edu/product/H08YUX-PDF-ENG> (cited Aug 15 2026)
  21. Schwartz E. et al., *More Students Use AI for Homework, and More Believe It Harms Critical Thinking*. RAND. Mar. 17 2026. Available from [https://www.rand.org/pubs/research\\_reports/RRA4742-1.html](https://www.rand.org/pubs/research_reports/RRA4742-1.html) (cited Aug 15 2026).
  22. Walton Family Foundation *Gen. Z Research*, Available from [gallup.com/analytics/651674/gen-z-research.aspx](https://www.gallup.com/analytics/651674/gen-z-research.aspx); Apr. 2025; Rubin, *Gen Z is Still Anxiously Using AI: Poll*. Axios, Apr. 8 2025, Available from <https://www.axios.com/2025/04/08/gen-z-artificial-intelligence-gallup-feelings>; Brenan M., *Gen Z's AI Adoption Steady, but Skepticism Climbs*, Apr. 8 2026. Available from <https://news.gallup.com/poll/708224/gen-adoption-steady-skepticism-climbs.aspx>; Adarlo S., *Usually, Young People Embrace New*

- Technology. Gen Z's Attitude Toward AI Should Worry the Entire Tech Industry.* Futurism. Apr 15 2026. Available from <https://futurism.com/artificial-intelligence/gen-z-attitude-ai> (all cited Aug 15 2026).
23. Pew Research Center, *How Teens Use and View AI*. Feb 24 2026. Available from [https://www.pewresearch.org/wp-content/uploads/sites/20/2026/02/PI\\_2026.02.24\\_Teens-and-AI\\_REPORT.pdf](https://www.pewresearch.org/wp-content/uploads/sites/20/2026/02/PI_2026.02.24_Teens-and-AI_REPORT.pdf) (cited Oct 12 2026). Survey of 1,458 US teens aged 13-17: 64% use AI chatbots and 54% use them for schoolwork; 59% say AI cheating is common at their school; among teens expecting a negative effect on society, 34% cite overreliance or loss of critical thinking or creativity, 25% job loss.
24. Common Sense Media, *Teens in the AI Era: Schoolwork and Skills That Matter*. Press release, Aug 18 2026. Available from <https://www.common Sense Media.org/press-releases/teens-who-use-ai-for-schoolwork-are-handing-off-critical-thinking-common-sense-media-research-finds> (cited Oct 12 2026). 70% of US teens use AI for schoolwork, nearly two-thirds of them say they offload critical thinking to it; 38% say AI leads them to generate fewer of their own ideas; only 27% say a teacher has explained what AI is and how it works.
25. McClain C. et al., *Young Adults in the U.S. Are Increasingly Wary of AI, Concerned It Will Take Jobs*. Pew Research Center. Aug 18 2026. Available from <https://www.pewresearch.org/short-reads/2026/08/18/young-adults-in-the-us-are-increasingly-wary-of-ai-concerned-it-will-take-jobs/> (cited Oct 12 2026). Survey of 3,488 US adults, June 22-28 2026: 55% of 18-29-year-olds are more concerned than excited about AI in daily life (31% in 2021, 47% in 2025); 11% more excited than concerned; 73% expect AI to lead to fewer jobs in the US over the next 20 years.
26. Pew Research Center, *How Opinions and Use of AI Differ by Age*. In *Americans and AI 2026*. June 17 2026. Available from <https://www.pewresearch.org/internet/2026/06/17/how-opinions-and-use-of-ai-differ-by-age/> (cited Oct 12 2026). Among US adults aged 18-29: 66% have used AI chatbots; 31% are extremely or very confident using them.
27. Eurostat, *64% of 16-24-Year-Olds Used AI in 2025*. Feb 10 2026. Available from <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20260210-1> (cited Oct 12 2026). 63.8% of EU 16-24-year-olds used generative AI tools in 2025, against 32.7% of 16-74-year-olds; 39.3% of 16-24-year-olds used them for formal education; highest in Greece (83.5%), lowest in Romania (44.1%).
28. Ofcom, *Passive Social Media Use, AI Companionship and Online Side Hustles: UK Adults' Media and Online Lives Revealed*. Apr 2 2026. Available from <https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-adults/passive-social-media-use-ai-companionship-and-online-side-hustles-uk-adults-media-and-online-lives-revealed> (cited Oct 12 2026). 79% of UK 16-24-year-olds use AI tools such as ChatGPT, Copilot or Gemini, against 54% of all adults.

29. Silver L. et al., *What Do Young People Around the World Think About AI?* Pew Research Center. Sept 17 2026. Available from <https://www.pewresearch.org/global/2026/09/17/what-do-young-people-around-the-world-think-about-ai/> (cited Oct 12 2026). Survey of 42,151 adults in 36 countries, Feb-May 2026: in the US, the share of 18-34-year-olds more concerned than excited about AI rose from 40% (2024) to 55%; in Sweden, among under-35s, from 24% (2025) to 45% (2026).
30. Ministry of Education Singapore, *Committee of Supply 2026: Announcements*. Mar 3 2026, last updated July 30 2026. Available from <https://cos.moe.gov.sg/2026-announcements/> (cited Oct 12 2026)
31. National University of Singapore, *Study Finds Over Half of 8-Year-Olds in Singapore Have Used AI; Most Popular Tool Is ChatGPT*. May 25 2026. Available from <https://www.nus.edu.sg/thought-leadership/study-finds-over-half-of-8-year-olds-in-singapore-have-used-ai--most-popular-tool-is-chatgpt> (cited Oct 12 2026). National study of nearly 3,000 children led by Prof. Jean Yeung; AI use rises to more than nine in ten by age 13.
32. BCS, The Chartered Institute for IT, *Reformed School Curriculum Will Feature AI, Media Literacy and a New Computing GCSE*. Nov 5 2025. Available from <https://www.bcs.org/articles-opinion-and-research/reformed-school-curriculum-will-feature-ai-media-literacy-and-a-new-computing-gcse/> (cited Oct 12 2026). On the UK Government's response to the Curriculum and Assessment Review.
33. Cuatrecasas, *Digital Omnibus on AI: How Does It Impact Employment Relations?* July 25 2026. Available from <https://www.cuatrecasas.com/en/global/labor-and-employment/art/digital-omnibus-on-ai-how-does-it-impact-employment-relations>; Van Eecke P., Regenhardt B., *Digital AI Omnibus Delays Key Deadlines, Introduces New Rules*. Cooley. Aug 3 2026. Available from <https://cdp.cooley.com/digital-ai-omnibus-delays-key-deadlines-introduces-new-rules/>; Interface, *An Operational Framework for AI Literacy in the Workplace*. Policy Brief. Aug 25 2026. Available from <https://www.interface-eu.org/index.php/publications/ai-literacy-workplace-framework> (all cited Oct 12 2026). The Digital Omnibus on AI entered into force on July 27 2026.
34. Gurley, B., *Running Down a Dream: How to Thrive in a Career You Actually Love*. Crown Currency, New York, 2026. ISBN 978-0-593-79966-6 (cited Aug 15 2026)
35. Gallup (2026). *State of the Global Workplace Report*. Available from <https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx> (cited Aug 15 2026)
36. British Brief. June 10 2026. *Record Number Think University Not Worth It*. Available from <https://britbrief.co.uk/education/universities/record-number-think-university-not-worth-it.html>. The first net-negative reading since 2005 and roughly double the 2018 share; confidence in financial returns at a record-low 36%; Isabel Berwick. *Is University Worth It?* Financial Times, July

- 2 2026, Available from <https://www.ft.com/content/f99266d1-f6fb-4183-bea1-9eecf9340fc5?syn-25a6b1a6=1> (all cited Aug 15 2026).
37. UNESCO (2026). *Higher Education Global Trends Report*. May 2026. Available from <https://www.unesco.org/en/articles/number-students-higher-education-more-doubled-20-years-inequalities-remain> (cited Aug 15 2026)
  38. Bouchrika I., *AI Bachelor's Degrees with Entrepreneurship Focus*, Aug 11 2026. Available from <https://research.com/online-degrees/artificial-intelligence/ai-bachelors-degrees-with-entrepreneurship-focus> (cited Aug 15 2026)
  39. *Teaching Entrepreneurship* (2026). Step-by-Step AI Lesson Plan for Entrepreneurship Students in 2026. July 23 2026. Available from <https://teachingentrepreneurship.org/ai-lesson-plan-for-entrepreneurship-students-in-2026/> (cited Aug 15 2026).
  40. Poth R. D., *From Idea to Impact: How AI Is Fueling the Next Generation of Student Entrepreneurs*. Getting Smart, July 7, 2025. Available from [dea-to-impact-how-ai-is-fueling-the-next-generation-of-student-entrepreneurs/](https://www.gettingsmart.com/2025/07/07/from-idea-to-impact-how-ai-is-fueling-the-next-generation-of-student-entrepreneurs/); Columbia Business School (2026), *AI Startup Challenge 2026: Applied AI and the Future of Business*. Available from <https://business.columbia.edu/ai-startup-challenge> (cited Aug 15 2026).
  41. Merida A. L. et al., *It's about time: The timing of entrepreneurial experience and the career dynamics of university graduates*. Research Science Direct. Jan 2021. Available from <https://www.sciencedirect.com/science/article/abs/pii/S0048733320302109?via=ihub> (cited Aug 15 2026)
  42. Botelho T. L. et al., *The Evaluation of Founder Failure and Success by Hiring Firms: A Field Experiment*. Organization Science, 2023. Available from [https://econpapers.repec.org/article/inmorsc/v\\_3a34\\_3ay\\_3a2023\\_3ai\\_3a1\\_3ap\\_3a484-508.htm](https://econpapers.repec.org/article/inmorsc/v_3a34_3ay_3a2023_3ai_3a1_3ap_3a484-508.htm) (cited Aug 15 2026)
  43. Selin Z. et al., *The Influence of Entrepreneurial Policy on Entrepreneurial Willingness of Students*, Frontiers, Educational Psychology, July 29 2021, Available from <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.592545/full> (cited Aug 15 2026)
  44. Greaves L., *What Do Graduates Do?* Prospects Luminate. Nov. 2025. Available from: <https://luminate.prospects.ac.uk/what-do-graduates-do> (cited Aug 15 2026).
  45. Sandoval F., *If Entry-Level Jobs Are Vanishing, Where Will the Future Executives Come From?* Ivy Exec, Aug. 12 2025, Available from <https://ivyexec.com/career-advice/2025/if-entry-level-jobs-are-vanishing-where-will-the-future-executives-come-from/>; Zhu K., *Charted: The Decline of US Software Developer Jobs*, Visual Capitalist. Mar. 14 2025. Available from <https://www.visualcapitalist.com/charted-the-decline-of-u-s-software-developer-jobs/> *The Junior Developer Is Going Extinct. What the Stanford Data Actually Shows*. Substack. Feb. 18 2026. Available from

- <https://mothasa.substack.com/p/the-junior-developer-is-going-extinct>; Bernard B., *Experience Requirements Have Tightened Amid the Tech Hiring Freeze*, Hiring Lab, July 30 2025, Available from <https://hiringlab.indeed.com/2025/07/30/experience-requirements-have-tightened-amid-the-tech-hiring-freeze/> (all cited Aug 15 2026)
46. Sandle T., *After Massive Tech Sector Layoffs in 2025, What's in Store for 2026?* Digital Journal. Jan. 8 2026. Available from <https://www.digitaljournal.com/article/after-massive-tech-sector-layoffs-in-2025-whats-in-store-for-2025/>; Muse D. et al., *Tech Layoffs: A 2025 Timeline*, Computerworld. July 23 2026, Available from <https://www.computerworld.com/article/3816579/tech-layoffs-this-year-a-timeline.html> (both cited Aug 15 2026)
47. Zeff M., *Microsoft CEO Says Up to 30% of the Company's Code Was Written by AI*. TechCrunch. Available from <https://techcrunch.com/2025/04/29/microsoft-ceo-says-up-to-30-of-the-companys-code-was-written-by-ai/> (cited Aug 15 2026)
48. *It's Bad Time To Be a Graduate*, Financial Times, July 6 2025, Available from <https://www.ft.com/content/002a0943-f977-44bd-bc78-957c877dfed1?syn-25a6b1a6=1> ; Staton B., *Five University Leavers On What It Really Takes To Land a Graduate Job*, Financial Times. May 28 2026. Available from <https://www.ft.com/content/16e2d97b-29e0-4a16-805e-f79ec2b15a5f?syn-25a6b1a6=1>; Staton B., *Why Recruiters Are Making Interviews AI Free Zones*, Available from <https://www.ft.com/content/bcc3becc-859e-4628-b6a7-b4788e6f20a6?syn-25a6b1a6=1> (all cited Aug 15 2026)
49. *7 in 10 Companies will Use AI in the Hiring Process in 2025, Despite Most Saying It's Biased*. Oct 22 2024. Available from <https://www.resumebuilder.com/7-in-10-companies-will-use-ai-in-the-hiring-process-in-2025-despite-most-saying-its-biased/> (cited Aug 15 2026); Insight Global's *2025 AI in Hiring Report*. Available from <https://insightglobal.com/2025-ai-in-hiring-report/> (both cited Aug 15 2026).
50. Kissin, E. *McKinsey Challenges Graduates to Use AI Chatbot in Recruitment Overhaul*. Financial Times, Jan. 14 2026. Available from <https://www.ft.com/content/de7855f0-f586-4708-a8ed-f0458eb25586> (cited Aug 15 2026)
51. Brynjolfsson E., Chandar B., Chen R., *No Widespread Displacement, but the AI Employment Gap for Young Workers Has Widened to 19%*. Stanford Digital Economy Lab. Aug 12 2026. Available from <https://digitaleconomy.stanford.edu/news/canariesaug26/> (cited Oct 12 2026). Updated analysis of ADP payroll data: employment of 22-25-year-olds in highly AI-exposed occupations is about 19% below its counterfactual as of June 2026 (15% at the July 2025 data vintage); the adjustment comes mainly through reduced hiring; the patterns are descriptive, not causal.
52. Original Source: Indeed Hiring Lab, *2026 Mid-Year UK Jobs & Hiring Trends Report*. Aug 2026. Used source: Onumonu M., *Graduate Job Postings at*

- Lowest Level Since Pandemic - Indeed*. Workplace Journal. Aug 3 2026. Available from <https://workplacejournal.co.uk/2026/08/graduate-job-postings-at-lowest-level-since-pandemic-indeed/> (cited Oct 12 2026). UK graduate postings around 7% lower year on year and at their lowest since 2020; 9.4% of UK job postings mentioned AI at the end of June 2026, a record high.
53. Original Source: International Labour Organization (ILO), *Global Employment Trends for Youth 2026*. Aug 11 2026. Used source: *World Youth Unemployment Rate Rose Last Year: UN*. CTV News. Aug 11 2026. Available from <https://www.ctvnews.ca/world/article/world-youth-unemployment-rate-rose-last-year-un/> (cited Oct 12 2026). Global youth unemployment rose to 12.4% in 2025; entry-level office and administrative jobs are declining, but the ILO considers it too early to attribute this to AI.
  54. Elements of AI, Available from <https://www.elementsofai.com/> (cited Aug 15 2026)
  55. Day of AI, RAISE Initiative. Available from <https://raise.mit.edu/resources/curricula/day-of-ai/> (cited Aug 15 2026)
  56. AI4All, Available from <https://ai-4-all.org/> (cited Aug 15 2026)
  57. Machine Learning For Kids, Available from <https://machinelearningforkids.co.uk/>; Experience AI with Google DeepMind. Available from <https://experience-ai.org/en/> (both cited Aug 15 2026)
  58. Lauterbach, A. *Romy, Roby and the Secrets of Sleep*. Available from [www.romyandroby.ai](http://www.romyandroby.ai) and on Amazon; AI Snacks on Roby&Roby, on all podcasting platforms. Instagram: @romyandroby (cited Aug 15 2026). Educational resources available on [www.romyandroby.ai](http://www.romyandroby.ai) on November 30th. AI Literacy Course available from December 24th 2026.
  59. Olsen A., et al., *Identification of Dual-Purpose Therapeutic Targets Implicated in Aging and Glioblastoma Multiforme Using PandaOmics – an AI-enabled Biological Target Discovery Platform*. Impact Journals: Aging. Apr 26 2023. Available from <https://pmc.ncbi.nlm.nih.gov/articles/PMC10188351/> (cited Aug 15 2026)
  60. Girls in Quantum. Available from <https://www.girlsinquantum.com/> (cited Aug 15 2026)
  61. <https://www.qubitbyqubit.org/> (cited Aug 15 2026)
  62. Druga S. et al., *Family as a Third Space for AI Literacies: How Do Children and Parents Learn about AI Together*, Digital Library, Apr 29 2022, Available from <https://dl.acm.org/doi/10.1145/3491102.3502031>; Xie J. et al., *Building AI Literacy at Home: How Families Navigate Children's Self-Directed Learning with AI*, available from <https://arxiv.org/html/2510.24070v1>; Tseng T. et al., *Collaborative Machine Learning Model Building with Families Using Co-ML*, June 14 2023, Available from <https://arxiv.org/pdf/2304.05444>; Li Y. et al., *Learning Together: AI-Mediated Support for Parental Involvement in Everyday Learning*, Computer Science. Available from <https://arxiv.org/abs/2510.20123> (all cited Aug 15 2026)

## Section 2: AI Literacy in Teachers

The people preparing our children for the future - teachers - are using AI without properly understanding how AI and generative AI technologies work. They are employed by institutions that mostly have no plan around AI Literacy and AI Fluency, and they live in countries whose governments have no strategic vision on how to make AI work for the economy, society and education. There are few exceptions to learn from - and the best news is that the problem is solvable if there is genuine willingness to address it. Every single teacher can move a needle for his/her classroom, even if fixing national policies won't happen any time soon.

We identified several themes describing the state of AI Literacy in schools:

First, *usage is near-universal; understanding is not*. In the edutAlnment survey of teachers (n=131 full responses from a dataset of 1,149),<sup>1</sup> 86% use ChatGPT or a comparable large language model — yet only 11% describe themselves as having advanced knowledge of AI, and fewer than 2% have ever built a tool with the help of AI/genAI. We call it the *practitioner paradox*: teachers have absorbed the consumer-facing layer of AI without the conceptual grounding to judge how these tools work, or to teach students how to relate to them critically.

Second, AI amplifies skills only in those who already have them. *Cognitive debt risk* is real in students (and adults). Teachers lack insights how to teach beneficial use of AI, or use AI as an infrastructure for personalized learning that would improve skills across different subjects – from mathematics and science to language and writing.<sup>2, 3</sup>

Third, *teachers fear that students use AI for cheating*, not as a tool to enhance existing abilities. Seventy percent of our respondents named students using AI to cheat as their top concern, far ahead of lost human connection (52.7%) and teacher replacement (46.4%).<sup>1</sup>

Fourth, the *institutional and political void* is the most consequential finding. Only 34.5% of educators agree their school provides adequate support for learning about AI, despite 78.2% agreeing AI can help them teach more effectively, and 65.5% believing AI technologies will fundamentally change their role within five years.<sup>1</sup> The OECD documents the same failure internationally.<sup>4</sup> In the absence of policy, teachers are setting the rules themselves.<sup>5</sup> So far, there is no infrastructure to educate teaching workforce across the globe.<sup>6</sup> There are pilots and individual initiatives to learn from, either shaped by AI vendors such as Google<sup>7</sup> or OpenAI and Anthropic<sup>8</sup> (not focused on AI Literacy, but on AI Fluency), or led by enthusiasts.<sup>9, 10</sup> Funding to remove the cost barrier is starting to appear.<sup>11</sup>

Fifth, *the main barrier to learn about AI properly is cost and time*. Teachers named cost (62.4%) and lack of time (54.1%) far above everything else, and two-thirds prefer short, self-paced modules of 15–30 minutes.<sup>1</sup>

Sixth, *a handful of governments have a vision to learn from*. Beijing has made AI education compulsory from age six;<sup>12</sup> the UAE mandates AI from kindergarten through Grade 12 with ethics built in from the start;<sup>13</sup> Chile embeds AI into initial teacher training but has no operational plan to scale to the existing workforce despite availability of content to do so. AI education becomes an additional thing to do on teachers' schedule, after their usual day of teaching, correcting the homework and doing administrative tasks is done.<sup>14</sup>

Seventh, *AI in education done well* exists already, as a scalable infrastructure across multiple subjects, and as individual projects. Alpha School runs AI underneath the learning as an adaptive tutor (their technology is not LLM-based).<sup>15</sup> Michael Leggs built an AI that coaches students through writing an essay instead of writing it for them.<sup>9</sup> The UAE's traffic-light rule gives every classroom at the Dubai British School Jumeirah Park a common language for when the machine helps and when it must be set aside.<sup>16</sup>

Eighth, *robot-proofing our children is a task for parents and teachers alike* - and the goal is to build cyborgs, not spectators or mindless consumers. In her new book *Robot-Proof: When Machines Have All the Answers, Build Better People*, the neuroscientist Vivienne Ming distinguishes three kinds of AI users: automators, who copy and paste; validators, who seek confirmation of what they already believe; and cyborgs, who spar with the machine and outperform the best humans and the best AI alike.<sup>17</sup> Raising cyborgs means giving children durable skills to thrive first — mathematics as structured reasoning, writing as thinking made visible, self-awareness, entrepreneurship, and curiosity - and then educating them about how AI can enhance those abilities, never replace building them. The edutAlnment whitepaper *AI Literacy in Young People and Students* shows that students are asking to be taught, not protected<sup>18</sup> - and that bans,<sup>19</sup> warnings, and detection theater destroy the trust on which learning depends. This work cannot be delegated to schools alone: families are a third space for AI literacy, and children whose parents engage in joint exploration learn to think critically about the technology rather than merely consume it.<sup>18</sup>

In the following pages, we examine how educators use AI, what they understand about AI and genAI technologies, and what they fear most about applying genAI in classroom. We review the institutional and political void in which they work, and the nascent policies providing hope for a future with improved AI Literacy among national populations of educators across public and private institutions. As in previous segments, we follow up with recommendations built on collected data, 100 interviews with AI developers, practitioners and educators, and several secondary sources.

## **The Practitioner Paradox: Teachers Between Use and Understanding**

In May 2026, edutAlnment GmbH surveyed educators around the globe. Of 1,149 survey visits, 131 people provided full answers - most of them based in the United States, and all holding a Master's or a professional degree.<sup>1</sup> The survey was sent out internationally, but mostly US-based participants decided to contribute. *This lack of wider participation is a data point in its own right.*

Anna Baron, a US student who interviewed seven teachers at her school for the edutAlnment project on AI Literacy, documented educators' anxiety with particular clarity.<sup>20</sup> Of the seven she approached, only four replied - a silence she read not as indifference but as evidence of how "controversial" AI remains: *"Most teachers are unsure how to publicly discuss it, and many are reluctant to take a position at all."*

The disciplinary divide Baron observed maps precisely onto the broader literature. Her humanities teachers were almost uniformly sceptical. A history teacher stated flatly: "I am not in favor of it," stressing the importance of handwriting and arguing that teachers cannot demand authentic work from students while relying on AI themselves. An English teacher offered a memorable analogy: AI was like "a robot lifting weights for someone at the gym. While the task may get completed, the person doing it gains none of the strength." Her STEM teachers took a different view. A mathematics teacher argued that AI has *"huge potential when used correctly"* - specifically when students use it as a personalised tutor to ask for hints, clarify concepts and identify mistakes rather than to obtain answers directly. She proposed that schools should *"create frameworks that help students to ask questions and use AI as a tool to further their learning."* A fourth teacher reframed the discussion around accessibility: AI could *"bridge skill gaps and expand opportunities"* for students with writing difficulties, language barriers or limited access to academic support.<sup>20</sup>

The cohort surveyed by edutAlnment is not a group of early adopters or digital natives. Almost everyone falls into the 30-44 (47.8%) and 45-59 (44.4%) age bands; not a single respondent was under 30. More than three-quarters have taught for over eleven years, and 40.8% for over twenty. Most work in public schools (64.6%), with strong representation from Special Education (28.5%), Languages and Humanities (26.9%), and Mathematics and STEM (24.6%).<sup>1</sup>

This matters because it conditions everything else. These are mid- and late-career professionals whose identities were formed around craft - lesson design, the student relationship, assessment - that predates the AI era. They bring experience shaped by decades of human contact with learning.

Asked to self-assess, teachers cluster in the middle: 46.1% report a basic understanding of AI, 32.2% a good one (able to explain it to others), 11.3% an advanced command, and 7.8% have merely heard of it.<sup>1</sup> Eighty-six percent use ChatGPT or a comparable large language model - yet fewer than 2% have ever built or taught AI on a level that would have satisfied students. One respondent put the practitioner paradox plainly: *"How does it actually work? I use it but I don't understand what's happening underneath. That gap makes me uncomfortable."*<sup>1</sup>

The more revealing numbers concern the ecosystem beneath the products. Proprietary chatbots (ChatGPT, Claude) are familiar to 64.3% - but predictive analytics to only 31.3%, natural-language processing to 27.0%, adaptive learning systems to 19.1%, and open-weight models such as the Llama family to a mere 7.0%.<sup>1</sup> The pattern is clear: *teachers know the product, not the mechanism* and the foundational mathematics powering AI. Their picture of AI is therefore narrow.

Fluency asks only that the tool works. Literacy asks that the person understand, at some level, why and how.

Technology scepticism might be viewed as a constant feature in educators. Punya Mishra reminds us that each new educational technology is first condemned as cognitively damaging, then quietly rehabilitated: Google Search was once “the villain,” only to become, a decade later, the benign baseline against which large language models are now measured.<sup>21</sup>

The framing that organizes this paper comes from a word coined by Joseph Aoun, president of Northeastern University: *robot-proof*. In his 2017 book of that name, Aoun argued that education should cultivate what automation cannot replicate - creativity, critical thinking, ethical judgment, and the capacity to keep learning.<sup>22</sup> A society that equips its teachers with knowledge and understanding of AI shapes a generation that can participate in an informed dialogue about its use, about what constitutes responsible and ethical AI, and about how to enhance human skills and capabilities with it. As automation and AI shift traditional employment structures, and as demographic decline across Western societies makes increased levels of automation, robotics, and AI inevitable, it is imperative to give teachers the attention and training they deserve.

## Cognitive Debt

We have already discussed the risk of cognitive debt when young people (and adults) outsource their thinking to a machine. In *The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI*, Barbara Oakley, Michael Johnston, Ken-Zen Chen, Eulho Jung and Terrence Sejnowski name the structural problem at the heart of the debate: as external aids grow more capable, internal memory can atrophy.<sup>23</sup>

The finding that matters most for teachers is a paradox within the paradox: *AI amplifies skill only in those who already have it*. You can evaluate an output, catch an error, or ask a good question only if you have internalized enough knowledge yourself. Without this internalization, AI does not close the gap between strong and weak learners — it widens it. The authors call the goal *augmentation without atrophy*: technology can strengthen thinking, but only when the learner has first done the cognitive work rather than skipped it.<sup>23</sup>

The neuroscientist Vivienne Ming sees the same evidence. In her account, the developers who leaned on AI “got worse at coding... worse at understanding the problem, worse at strategic thinking,” because “when you offload your cognition into the machine, not only do you not learn while you’re doing it, your cognition atrophies.”<sup>17</sup> And the loss, she warns, is not confined to cognition: “if you offload your emotional stability, your trust, your social skills to a machine, then you have a serious problem.” Her inversion of the usual sales pitch is the useful design principle. We sell AI as relief - less effort, faster output - when the honest use is the opposite: *productive friction*, using the tool to make the work harder in exactly the ways that make the learner better. That is the difference between the student who hands the

question to the machine and copies the answer, and the one who uses the machine to sharpen a question worth answering.

## The Cheating Prism

Seventy percent of surveyed educators named students using AI to cheat as their leading concern. Nothing else came close: loss of human connection (52.7%), fear of teacher replacement (46.4%), and privacy (38.2%) all ranked well below.<sup>1</sup> We call it the *cheating prism*: AI reaches most classrooms filtered through academic integrity before any other consideration can arrive. One open response describes the essence of it: *“My biggest challenge is that students use it to complete assignments without learning anything. I’ve had to completely redesign my assessments.”*<sup>1</sup>

AI magnifies what is already there. Long before ChatGPT, cheating was endemic - in surveys of more than 70,000 US high school students between 2002 and 2015, 64% admitted cheating on a test, and Stanford's Challenge Success surveys have found for over fifteen years that 60 to 70 percent of students report at least one recent cheating behavior. When Stanford researchers Victor Lee and Denise Pope re-ran the survey across 40 high schools after ChatGPT's release, the number was statistically unchanged - and had even decreased slightly.<sup>24</sup> Students cheat, their data show, not because a tool exists but because they are disengaged, stressed, sleep-deprived, or do not feel they belong; the technology merely lowers the cost of a decision to cheat. The long-term stakes of that decision are documented: students who engage in academic dishonesty are significantly more likely to engage in dishonest behavior in the workplace - a moderate-to-strong correlation replicated across multiple studies and countries, extending to fraud, corruption tolerance, and professional misconduct.<sup>25</sup> This is what great teachers can communicate, and how: not “you might get caught,” which frames integrity as a detection problem and invites an arms race, but *“cheating is practicing to be someone you don’t want to become,”* and the gym-weights analogy from Baron’s English teacher makes the point in one sentence a teenager remembers.<sup>20</sup>

Interactive oral examinations make cheating difficult while making structured learning (and genuine understanding) visible. Research at Griffith University by Sotiriadou and colleagues, using scaffolded assessments culminating in interactive orals, found that they help prevent academic misconduct, and that the more relevant the assessment to real-world scenarios, the less likely students are to engage in misconduct - while simultaneously building professional identity, communication skills, and employability.<sup>26</sup> A 2024 systematic review confirms the pattern: oral formats were repeatedly introduced precisely in response to misconduct in written assessments, students report that contract cheating is difficult to impossible in an interactive oral, and - the finding teachers should hear - most students performed *better* orally than in writing, and prepared more thoroughly because they knew they would have to explain themselves.<sup>26</sup>

## The Institutional and Political Void

The most consequential finding in the survey is not about knowledge or fear. It is about support. Only 34.5% of educators agreed their school provides adequate support for learning about AI - even as 78.2% agreed AI can help them teach more effectively, and 65.5% believed it will fundamentally change their role within five years.<sup>1</sup>

Teachers can see the transformation coming and believe the technology can help — yet fewer than one in three feel prepared for it by their institution. One wrote: *“My school has no policy. Teachers are left to figure it out alone. Some ban it, some ignore it. There’s no coherent approach.”*<sup>1</sup> This is the education-sector version of a pattern visible across the workforce: capable people stuck in institutions that have not caught up.

The failure is documented internationally. The OECD’s Digital Education Outlook 2026 - the most comprehensive cross-national analysis to date - reports that a large share of teachers already use generative AI for work, while warning that overreliance can undermine “teacher autonomy and professionalism” and erode teachers’ own skills. Its recommendation is telling: a shift toward AI systems designed *with* teachers - which presumes teachers who understand those systems well enough to shape them.<sup>4</sup>

The deeper problem is structural. Where governments have no vision for their country’s future, they have no vision for its education. AI policy becomes a vacuum filled, classroom by classroom. WIRED’s reporting captured the US version: “Some districts have issued guidance. Others have thrown up their hands. In the absence of clear policy, teachers are setting the boundaries themselves.”<sup>5</sup>

## Private versus Public Schools

The divide that should worry us most runs between private and public education - and what it means for the future of the workforce. In the survey, most respondents of edutAlnment survey work in public schools, where support is thinnest.<sup>1</sup> At the other extreme sit well-funded private microschools building bespoke AI tutoring. The market is moving fast: the global market for AI in education, valued at \$7.05 billion in 2025, is projected to reach \$136.8 billion by 2035 - a nearly twenty-fold increase in a decade - though estimates vary, with other analysts projecting \$57-80 billion over similar horizons.<sup>27</sup> WIRED documents a Silicon Valley movement - Musk’s Ad Astra and its spin-offs, Alpha School, venture-funded platforms - animated by a “tech exit” philosophy of opting out of public systems.<sup>28</sup> Josh Cowen, an education-policy scholar, warns where that leads: *“You remove yourself from these big structural systems... and you just say, ‘You’re on your own.’ It works well for some people, but it doesn’t work for most people.”*<sup>29</sup>

Scaling best practices must not be limited to private systems. The smart question for any ministry or district is not where an approach comes from, but whether it is the best one available for building internalized knowledge - and then to fund it

accordingly. Public systems have borrowed from private innovation before: the International Baccalaureate - created for private international schools - is today taught mostly in public schools in the US and was adopted by Ecuador's ministry for its state system at national scale, and Montessori pedagogy runs in hundreds of public schools worldwide. The lesson from those transfers, and from Ecuador's uneven results, is the same one Chile teaches: borrow the idea and the design *and* build the operational support to ensure best practices are truly implemented.

### **Case Study - Alpha School: AI Underneath, Humans on Top**

Alpha School, founded in 2014 by MacKenzie Price and now operating across US campuses under a “Two Hours Learning” model, has students complete core academics - maths, reading, language, science - in about two focused morning hours, then spend afternoons on projects and life skills.<sup>15</sup> On the NWEA MAP assessment, Alpha reports placing its classes around the 99th percentile across most grades and subjects, and a measured learning rate roughly 2.3× the predicted pace - an attempt to deliver, at scale, Benjamin Bloom’s famous “2 Sigma” tutoring effect. What makes Alpha instructive is not the headline numbers, but three design choices that run against the naive “give every child a chatbot” model:

*No open chatbot for students.* Price is explicit that a chatbot interface “generally... is going to be used to cheat.” AI runs underneath as an adaptive tutor and diagnostic engine - measuring pace and accuracy, curating the next lesson - not as an answer machine on top.

*Mastery before amplification.* The model rests on mastery-based progression, spaced repetition, and productive struggle. Children build the schema first; AI accelerates the practice loop.

*Teachers become guides.* Alpha calls its teachers “guides.” The AI handles content delivery; the human handles what a machine cannot - motivation, mentorship, and the judgment to intervene.

Alpha tuition is around \$40,000 (similar price tag as in European IB schools or national private boarding schools).

### **Obstacles to Implement AI Literacy**

Asked what prevents them from getting AI-literacy training, teachers named cost (62.4%) and lack of time (54.1%) far above everything else. Lack of school support registered at 20.2%; only 15.6% were unconvinced of the relevance.<sup>1</sup> Teachers who lack AI literacy are not indifferent - they are structurally blocked. Forty percent describe themselves as extremely or very interested in training, another 27.5% as moderately interested.<sup>1</sup>

The format they want confirms it. Sixty-five percent prefer short, self-paced online modules of 15-30 minutes - far ahead of live webinars (37%) or in-person workshops

(35%).<sup>1</sup> A teacher who cannot carve out a full day can manage a module between lessons. These are not stylistic preferences; they are constraints translated into requirements.

The encouraging news is that the money to remove the cost barrier is starting to appear. In January 2026, Google.org gave California State University, Fullerton, \$500,000 to train teachers in AI literacy (unfortunately, Google pushes more for AI fluency than for literacy itself); the non-profit aiEDU launched grants of \$25,000–\$50,000 for educator capacity in under-served communities; the National Science Foundation committed roughly \$11 million to K-12 AI teacher training; and the US Department of Labor opened workforce grants with an explicit AI-literacy emphasis.<sup>7</sup>

## **Where Governments Have a Vision**

A handful of governments have decided that AI in education is a national question. They are worth studying - for what they get right, and for what their haste reveals.

### ***China***

Beijing became the first provincial-level region in China to make AI education compulsory: from September 2025, more than 1,400 primary and secondary schools must provide at least eight class hours of AI per year, starting at age six. The curriculum is progressive by design - early grades explore everyday AI and privacy; middle grades build simple projects; upper grades meet models such as decision trees and neural networks; high schools work on intelligent agents. Hangzhou went further, mandating at least ten hours a year. The framing is explicitly strategic: AI literacy as part of the national talent policy, backed by very large education spending.<sup>12</sup>

### ***The United Arab Emirates***

In May 2025, Sheikh Mohammed bin Rashid Al Maktoum announced that AI would become a mandatory subject in every UAE public school from kindergarten through Grade 12, rolled out from the 2025-26 year across roughly 400,000 students with about 1,000 specially trained teachers. The curriculum is age-differentiated — storytelling and play for four-year-olds, system design and bias for ages 10–14, command engineering and real-world simulation in secondary school — and, notably, ethics is built in from the start. One school’s “traffic-light” system captures the pedagogy: red means no AI (the child’s own work), amber means teacher-directed use, green means supervised exploration. Even here, the guiding principle is that the teacher is irreplaceable — as one head of digital strategy put it, an AI tool cannot look at a child’s face and see that something is wrong.<sup>13</sup>

## Chile

Chile is the instructive case — real ambition with suboptimal execution. Its Política Nacional de Inteligencia Artificial 2021–2030 names education as a target sector, and in 2025 the Ministry of Science, with CENIA, Kodea and the UMCE, launched “Future Teachers in AI,” embedding AI into initial teacher training. A pilot of about 200 pedagogy students scaled toward a program intended to train thousands more, on an explicit multiplier logic: each teacher trained reaches many classrooms. The concept operationalizes the OPED framework at UC Chile, adding modules including “AI as an Academic Tutor” and “AI and Academic Integrity.”<sup>14</sup> Yet for many teachers the content arrived as an *additional burden*, with no plan for how to fit it into already-full workloads — good material, no operational design. That gap between having content and having an operating plan is exactly what makes best intentions fail.

## And Others

The list is growing. Egypt has made AI and programming compulsory for first-year secondary students; India’s CBSE has offered AI as an elective since 2019 and is weighing a wider rollout from Class 6; and England’s Tony Blair Institute has named the “AI-literacy gap” among teachers as a significant obstacle to innovation.<sup>30</sup> Three further cases deserve closer study, because each answers a question the others cannot: Estonia shows how a small state can adopt frontier AI on its own terms; Singapore shows guardrails designed as pedagogy; and South Korea shows how even a well-funded central mandate fails when it outruns its teachers.<sup>30</sup>

## Estonia

Estonia’s education minister frames the stakes bluntly: nations that equip teachers and students with AI literacy “will secure a decisive advantage.”<sup>31</sup> The country’s AI Leap (TI-Hüpe) — the explicit sequel to its 1996 Tiger Leap computerization program — acts on that conviction, and it is the clearest case of a public system adopting tools from the generative-AI foundational labs rather than building its own: from September 2025, 20,000 upper-secondary students and thousands of teachers received free access to AI learning tools negotiated with OpenAI and Anthropic, expanding toward roughly 58,000 students and 6,700 teachers by 2027 in a country of 1.3 million.<sup>31</sup> Student personal data is firewalled under GDPR and excluded from the vendors’ training sets; the program requires data minimisation and anonymisation and runs compliance checks on all contracted vendors; Estonia’s X-Road infrastructure keeps every data exchange signed, encrypted, and logged; and researchers at the University of Tartu worked with OpenAI to modify the tool so it behaves as a Socratic tutor that prompts with questions rather than handing over answers.<sup>31</sup> Three design choices complete the picture: multi-vendor procurement to avoid lock-in, teacher onboarding *before* student access, and a longitudinal study with Tartu and Stanford measuring cognition rather than logins.<sup>31</sup> Estonia works with

the big labs, in short, because the guardrails, the data pipeline, and the evidence base remain sovereign.

### ***Singapore: Guardrails as Pedagogy***

Singapore reaches a similar destination through a state-built platform. Its Ministry of Education delivers AI features through the national Student Learning Space under an AI-in-Education Framework that builds on the country's Model AI Governance Framework, with the explicit goal of minimising “cognitive offloading and dependence on AI.”<sup>32</sup> The learning assistant LEA is teacher-activated, limits the number of student interactions, asks guiding questions, and “will not provide spoon-fed answers”; every student completes a basic AI-literacy module covering how AI works, its benefits and risks; and human-in-the-loop moderation is mandated by GovTech-MOE guidelines.<sup>32</sup> By 2026, AI training is to be offered to teachers at all levels, including those still in training.<sup>33</sup>

### ***South Korea***

South Korea shows what happens when a centralized system moves fast without “carrying” its teachers. AI digital textbooks launched nationally in March 2025 for mathematics, English, and informatics, backed by \$760 million for teacher training over three years — and were stripped of official textbook status after a single semester, following complaints from teachers, students, and parents about factual inaccuracies, data-privacy risks, screen time, and *increased* teacher workload; adoption collapsed from 37% to 19% in one semester.<sup>34</sup> The failure was not the ambition but the failed process — a rushed rollout, minimal piloting, and a change of government — and it delivers the same lesson as Chile from the opposite direction: content and money without a stable operational plan for teachers doesn't pay into the democratization of AI literacy in a population.<sup>34</sup>

### **Centralized vs. Federal Design for AI Literacy Initiatives**

Federal state design, whatever its other virtues, is structurally damaging for AI-literacy initiatives, because the one thing such initiatives need - a single decision, funded and sequenced - is the one thing federations cannot guarantee. In the United States, education is constitutionally a state and local matter: Washington can convene task forces and issue executive orders, but the result on the ground is a patchwork of non-binding state guidance documents, district-by-district improvisation, and the teacher-level rule-setting our survey documents.<sup>5</sup> Switzerland is the same design in miniature: twenty-six cantons hold educational sovereignty, coordination runs through consensus bodies, and the current curricula's media-and-informatics strands predate generative AI - so updating them requires intercantonal agreement on a technology that changes faster than any such agreement can be reached.

Honesty requires the other half of the equation: the same federal design that blocks a national AI curriculum also enables policy experimentation across fifty states or twenty-six cantons, adapts schooling to local languages and needs, and - as South Korea's whiplash shows — prevents any single government from imposing a badly designed system on every classroom at once, or ripping a good one out. The federal task is therefore not about one mandate, but shared frameworks, funded interoperability, and fast mechanisms for scaling best practices. Federalism requires AI-literate leaders across the country, who can agree on what best practices might be, and what kind of society and economy AI technologies should support.

## **Recommendations: AI Literacy in Teachers**

Everything above points to one conclusion: there is no scalable system to address AI Literacy in education across geographies and school systems. There are teachers using AI, a few governments mandating it, and scattered pilots funding it. Building the system means answering two questions together: how to educate the young, and how to educate their teachers. Neither works without the other. This section sets out five recommendations, circling around the major theme: robot-proofing the next generation is the most strategic task for everyone in the society.

### *1. Start from a Vision, Not an AI Tool*

A country that knows where it is going can decide how to govern a transformational technology, and what kind of AI technology to implement. Governance is the operational expression of strategy, and education is where that strategy either takes root or fails. *But the same principle scales down - and it does not require national policy.* A school is a small universe: a principal who articulates what the school wants its graduates to be able to do can govern AI within those walls - choosing tools, setting the traffic-light rules, protecting the human skills - years before any national framework arrives. A classroom is smaller still, and a single teacher who has decided what learning must be protected and what the machine may amplify is, within those four walls, a complete AI-governance regime. Nothing in the institutional and political void prevents a principal or a teacher from stepping up and leading; the void is precisely what makes their bottom-up leadership decisive. The evidence in this paper points the same way: in the absence of policy, teachers are already setting the rules - the only question is whether they do it by improvisation or by intention. Vision, strategy, and governance are not privileges of scale. They are decisions, and a decision fits inside a classroom.

The following five questions can open up a dialogue within a single school, or even a classroom. They require no technical knowledge, work in any subject, and are addressed to students as partners rather than suspects - each one builds the shared understanding of AI in learning:

*“What should you be able to do without AI - and why does that matter?”* The foundational question. It forces the class - and the teacher - to name the skills worth protecting: mental arithmetic, a first draft, an argument held in the head - and to justify them, turning any AI usage rule from an imposed ban into a shared commitment.

*“Where has AI already helped you learn something - and where did it do the work instead of you?”* This legitimizes honesty about actual use, which is already happening, and hands students the central distinction - amplification versus substitution - as a lens they apply to their own behavior rather than a rule applied against them.

*“How do you know when the AI is wrong - and what would you need to know to identify mistakes?”* The literacy question. It surfaces hallucination and bias, and makes the case for internalized knowledge from the student’s side: you can only supervise a machine in a subject you understand.

*“If AI can produce the assignment, what should the assignment become?”* Asked openly, this turns students from adversaries of assessment into co-designers of it - and reliably generates the answers the research endorses: explain it aloud, defend it in dialogue, show the process, do it live. It is the cheating conversation, held before there is anything to catch.

*“What is one thing in our subject you would explore if a tireless, endlessly patient tutor were available - and what would you want it never to do?”* The forward-looking question. It surfaces genuine learning ambitions AI could serve, and lets students define the guardrails (e.g., “never just give me the answer.”)

The questions sequence naturally: the first three build the shared understanding, the fourth converts it into assessment practice, and the fifth opens the door to introducing AI applications into learning - with the class’s own rules already on the table. *What should follow the dialogue is an AI literacy course* - for the students and their teachers alike - addressing how AI technologies actually work and why: how a large language model differs from human thinking, why a next-token predictor produces fluent text and confident mistakes, what the machine can and cannot know, the meaning of synthetic data, and how to use AI so that it amplifies internalized knowledge rather than replaces building it. The dialogue creates the willingness; the course supplies the understanding.

## *2. Protect the Human Skills First*

The goal of a robot-proof education is not to produce children who can prompt a chatbot. It is to build the capacities a machine cannot: mathematics as structured

reasoning, creative and analytical writing as thinking made visible, and the broad muscles of creativity and problem-solving. Vivienne Ming argues that the only durable job description is *creative, adaptive problem explorer*, and that what predicts positive life outcomes is meta-learning, not memorized skills soon to be automated.<sup>17</sup> Her warning - the *Informational-Exploration Paradox* - is that automating the routine tends to produce *more* routine, not more creativity, unless we deliberately build both AI and people for exploration; the durable human work lies in “ill-posed problems” - messy challenges with no single right answer, which reward metacognition, creativity, curiosity, and socio-emotional skill. Ming states the classroom rule bluntly: “the golden rule of AI tutors - if they ever give students the answer, they never learn anything.”<sup>17</sup> The classroom implication is direct: use the time AI frees to deepen human skills, not to offload them.

While discussing AI Literacy, we can't forget two areas of knowledge that greatly contribute to enhancing human critical thinking and the ability to communicate. At edutainment, we encourage schools and universities in any country to compulsorily teach students of all age groups two subjects - (*creative*) *writing* and *applied mathematics*.

The first stimulates critical thinking and encourages students to separate original ideas from cheap prompting. As the linguist Naomi Baron notes, writers have talked about the role of writing in thinking for a century - Flannery O'Connor's “I write because I don't know what I think until I read what I say,” echoed by Forster, Koestler, Didion, and (by attribution) Faulkner; Stephen King's compact version, “I write to find out what I think,” appears in “The Colorado Kid” (2005). If writing helps us think, Baron asks, what happens when we surrender the process to AI?<sup>35</sup>

We could go on to find historical evidence on the value of writing, but the point is clear. Any professional can benefit from the clarity of thought as it evolves through serious training in writing skills. Jennifer A. Doudna not just won the 2020 Nobel Prize in Chemistry but popularized her research while writing “A Crack in Creation.”<sup>36</sup> David Deutsch isn't just one of the best quantum physics on the planet. He educates us while writing “The Fabric of Reality”<sup>37</sup> and “The Beginning of Infinity.”<sup>38</sup>

On another hand, Teaching applied mathematics to everyone isn't just about popularizing numerical skills. It is about reducing fears and showing that something as abstract as equations might be fun and valuable. Reading statistics and visualizing data is one of the required skills across areas of science and professions. When discussing applied mathematics, we don't need to reinvent the wheel. The International Baccalaureate (IB) Diploma Programme has been offering a course, “Mathematics: Applications and Interpretations” (MAI) for quite some time now, greatly benefiting those students who seek future degrees in medicine, psychology, or the arts.

### *3. Use AI as Infrastructure, Not as an Answer Machine*

The right role for AI is underneath the learning, as infrastructure for personalized tutoring and diagnostics - the Alpha Schools pattern - not on top of it as an answer machine.<sup>13</sup> Two rules follow, and they are not contradictory. Do not use LLMs to write students' essays. Do build and adopt systems, like Michael Leggs's, that coach students through the work.<sup>14</sup> The measure of any classroom AI is simple: does it leave the student with more internalized capability, or less?

#### *Case Study: Michael Leggs and his AI Essay Tool*

Speaking on the AI Snacks podcast, the US educator Michael Leggs described building an AI that guides students through *how* to write an essay - prompting them to form a thesis, structure an argument, and revise - rather than writing the essay for them.<sup>9</sup> The AI plays here a role of a Socratic coach, not oracle. The student remains the active agent and keeps the cognitive work that builds the skill.

### *4. Make AI Literacy Compulsory - for Students and Teachers*

Both groups need a compulsory course, and at this point in time, it is largely the same course. It should teach how AI differs from human intelligence and why; why these systems make confident mistakes (a next-token predictor has no truth-seeking mechanism); and how to use AI ethically and well. For teachers, five components follow from the evidence in this paper:

- *How LLMs actually work versus how humans think* - enough to explain to a class why a model produces fluent text full of hallucinations, and is not, in any real sense, "thinking." This is the gap between 64% ChatGPT familiarity in edutAlnment survey and near-zero comprehension of the mechanism.<sup>1</sup>
- *AI risks, including cognitive debt* (explaining constraints due to fundamental architectures of AI systems, the concepts of human involvement and supervision, or lack of it, overreliance on AI, and catastrophic risks)
- *AI in economy and society*, including how open-weight models can be used across areas of knowledge, application of Machine Learning and genAI in disciplines that are addressed in schools (from arts to chemistry, from biology to design).
- *Equity and accessibility* - how AI can support learning differences, language barriers, and limited access to help.
- *Institutional and policy literacy* - the vocabulary to shape school AI policy rather than merely receive it.

And the delivery must fit the profession: short, self-paced, accessible, and free or heavily subsidized.<sup>1</sup>

### *5. Introduce AI by Age - a Staged System*

Children do not need the same relationship with AI at six and at sixteen.

*Ages 4–7 (early primary): no independent AI use.* Playful, unplugged concepts - how are machines different from people? - and the sheer joy of reading, counting, drawing, and telling stories by hand. Anastassia Lauterbach has interviewed several entrepreneurs developing technologies for early-childhood education that are disconnected from the Internet, providing a walled-garden, secure approach to enabling skills such as meditation, reading, and counting.<sup>39</sup>

*Ages 8–11 (upper primary): AI made visible, teacher-directed only.* Children see AI tutors and image tools demonstrated, learn that AI can be confidently wrong, and keep doing the core cognitive work - arithmetic, handwriting, first drafts - themselves. They learn about AI via edutainment concepts, e.g., edutainment Romy&Romy.

*Ages 12–14 (lower secondary): supervised, structured use.* Socratic tutoring systems (the Michael Leggs case study; Alpha School Examples), AI as a study partner that gives hints not answers, and explicit lessons on bias, hallucination, and ethics. The traffic-light rule governs every task. Students can further use edutainment concepts to learn about how AI works.

*Ages 15–18 (upper secondary): guided autonomy.* Students use AI for genuine work -research, coding, analysis - while learning to verify, cite, and interrogate it, and while being assessed on judgment, not recall. They graduate literate, not merely fluent. Students in this group should be able to build a digital second brain, enabling their research and document management and preparing them for the years after school graduation. In this group, an AI Literacy course is mandatory.

The through-line is constant: *at every age, the human builds the capability first. AI amplifies it afterwards.*

None of the changes described in this paper are easy - not at the level of a nation, not at the level of a school, and not in an individual classroom. Curricula resist revision, budgets resist reallocation, and habits - institutional and personal - resist everything. What the evidence collected here requires is not another mandate but individual leadership: ministers, principals, and teachers willing to add new skills to their own repertoire while safeguarding the value of human agency - the capacity of a learner to think, to struggle productively, and to own the work of their own mind. AI will keep improving whether education engages with it or not. The leaders who matter now are those who engage on their own terms: learning how the technology works, deciding what it may amplify and what it must never replace, and modelling for their students the most important lesson of the AI era - that the human remains the agent of the outcome.

## Sources for Chapter 3\_Section 2

1. edutAlnment Survey: *AI Literacy in Educators*. Primary survey instrument; 131 full responses from 1,149 survey visits, mostly US-based, education sector, Master's or professional degree holders. Analysis including respondent profile (Section 1), AI familiarity and tool usage (Section 2, Q8), preferred formats (Q15), sentiment, concerns and attitudes (Q16–Q19), interest and barriers (Q21, Q23), and open feedback (Section 7). Original survey available upon request.
2. Malek A. et al., *Most Teachers Receive No Formal Guidance on AI Use*, Gallup, Available from <https://news.gallup.com/poll/710534/teachers-receive-no-formal-guidance.aspx>. (cited Aug 15 2026). According to Gallup, six in ten teachers use AI for their work, just 18% report receiving any formal guidance from school administrators on how AI tools should be used; across ten tasks, about one-third (34%) receive no guidance at all and about half (48%) only informal guidance. Directly on the personalized-learning half: 69% of teachers report receiving absolutely no guidance on how to apply AI to one-on-one instruction or tutoring, and 58% receive none on using AI for grading and student feedback. Methodology: web survey of 2,069 US public K-12 teachers, Feb. 9–Mar. 2 2026, recruited from the RAND American Teacher Panel.
3. Doss C. et al., *AI Use in Schools Is Quickly Increasing but Guidance Lags Behind. Findings from the RAND Survey Panels*. Sept. 30 2025. Available from [https://www.rand.org/pubs/research\\_reports/RRA4180-1.html](https://www.rand.org/pubs/research_reports/RRA4180-1.html) (cited Aug 15 2026)
4. OECD, *Digital Education Outlook 2026*. Jan 19 2026. Available from [https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026\\_062a7394-en.html](https://www.oecd.org/en/publications/oecd-digital-education-outlook-2026_062a7394-en.html) (cited Aug 15 2026)
5. Barnett S., *Teachers Are Trying to Make AI Work for Them*. Wired. Aug 18 2025. Available from <https://www.wired.com/story/teachers-using-ai-schools/> (cited Aug 15 2026)
6. Miao et al., UNESCO, *AI Competency Framework for Teachers*. Update Jan 16 2026. Available from <https://www.unesco.org/en/articles/ai-competency-framework-teachers> (cited Aug 15 2026)
7. Phillips C., VP Education/Google, *Our Commitment to Make AI Training Available to All 6 Million US Educators*, Available from <https://blog.google/products-and-platforms/products/education/teacher-ai-literacy-training/>; Kingkade T., *Inside Google's AI Training for Teachers*, NBC News, June 9 2026, Available from

<https://www.nbcnews.com/news/education/google-ai-training-teachers-california-rcna348979> (both cited Aug 15 2026)

8. AFT Press Release, *AFT to Launch National Academy for AI Instruction with Microsoft, OpenAI, Anthropic and United Federation of Teachers*. July 8 2025. Available from <https://www.aft.org/press-release/aft-launch-national-academy-ai-instruction-microsoft-openai-anthropic-and-united> (cited Aug 15 2026)
9. Michael Legg's *Socratic Writing AI Coach*, AI Snacks with Romy&Roby, Podcast, #100, Transcript Available on Request, podcast will be released on Nov. 24 2026, (transcript cited Aug 15 2026)
10. Watkins M., *Our Winter AI Institute for Teachers*. Substack. Available from <https://marcwatkins.substack.com/p/our-winter-ai-institute-for-teachers> (cited Aug 15 2026)
11. \$500,000 Google.org grant to train teachers: *Google.org Funds Teacher and Faculty Training in AI Literacy*. Cal State Fullerton. Jan 21 2026. Available from <https://news.fullerton.edu/2026/01/google-org-funds-teacher-and-faculty-training-in-ai-literacy/>; \$25,000-50,000 for education in under-served communities: *aiEDU, AI Literacy Should Be a Basic Component of Every Student's Education*. Available from: <https://www.aiedu.org/>; Vilcariono J., *Data Literacy Skills Seen as Key for Success in the Age of AI*. Available from <https://www.edweek.org/technology> (all cited Aug. 15 2026). Further grants and sources exist.
12. Shuo Z., *Beijing Makes AI Education Compulsory in Public Schools*. Oct 21 2025. China Daily. Available from <https://www.chinadaily.com.cn/a/202510/21/WS68f6d9bea310f735438b601e.html>; Buntz B., *China Mandates Eight Hours of AI Teaching From Grade One; US Executive Order Offers Incentives Instead*. Apr 25 2026. Available from <https://www.rdworldonline.com/china-mandates-eight-hours-of-ai-teaching-from-grade-one-us-executive-order-offers-incentives-instead/> (both cited Aug 15 2026)
13. *UAE Schools To Teach AI Starting in Kindergarten*. NewsOnAir, Available from <https://newsonair.gov.in/uae-schools-to-teach-ai-starting-in-kindergarten/> (cited Aug 15 2026)
14. *Inicia Futuros Docentes en IA: Alianza Público-Privada Para Integrar La Inteligencia Artificial En Carreras de Pedagogía*. Mar 5 2026. Available from <https://www.minciencia.gob.cl/noticias/inicia-futuros-docentes-en-ia-alianza-publico-privada-para-integrar-la-inteligencia-artificial-en-carreras-de-pedagogia/> ; CENIA, *Alianza Entre Estado, Academia Y Sociedad Civil Busca Llevar La Inteligencia Artificial A La Formación Inicial Docente En Chile*. July 24 2026. Available from <https://cenia.cl/2026/07/24/alianza-entre-estado-academia-y-sociedad-civil-busca-llevar-la-inteligencia-artificial-a-la-formacion-inicial-docente-en-chile/> ; Mineduc *Lanza Guía Para Potenciar El Aprendizaje*

*Integrando el uso de la Inteligencia Artificial en Contextos Educativos*. Mar 12 2025. Available from <https://www.mineduc.cl/mineduc-lanza-guia-para-potenciar-el-aprendizaje-integrando-el-uso-de-la-inteligencia-artificial-en-contextos-educativos/>; OES, *Guía De Inteligencia Artificial Para Docentes en Chile*. Mar 17 2025. Available from <https://oes.fundacion-sm.org/eduforics/reimaginar-juntos-los-futuros/tecnologias-y-aprendizaje/guia-de-inteligencia-artificial-para-docentes-en-chile/>; edutainment analysis based on the interview with Elisa Torres Durney, Chilean-born founder of Girls in Quantum.

15. Labenz N. 2-Sigma in 2 Hours: How Alpha Schools are Using AI to Revolutionize Education. June 25 2025. Available from <https://www.cognitiverevolution.ai/2-sigma-in-2-hours-how-alpha-schools-are-using-ai-to-revolutionize-education/>; Alpha Schools Website available from <https://alpha.school/> (cited Aug 15 2026)
16. AI Goes to School in the United Arab Emirates. Update Apr 6 2026. Available from <https://courier.unesco.org/en/articles/ai-goes-school-united-arab-emirates> (cited Aug 15 2026)
17. Ming V., *Robot-Proof: When Machines Have All the Answers, Build Better People*. Wiley, 2026; Ming V. interviewed by A. Lauterbach, *AI is Simply the Forcing Function*, Sept. 2026 and Episodes 76-77 of the *AI Snacks with Romy&Roby* podcast (*Can Humans Compete With Machines with Dr. Vivienne Ming, Parts 1 and 2*).
18. Lauterbach A., *AI Literacy in Young People and Students*. Whitepaper. Available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Aug 15 2026). Update with new data – December 1st 2026.
19. Burns M. et al., *A New Direction for Students in an AI World: Prosper, Prepare, Protect*. Brookings. Jan 14 2026. Available from <https://www.brookings.edu/articles/a-new-direction-for-students-in-an-ai-world-prosper-prepare-protect/> (cited Aug 15 2026)
20. Anna Baron, *What My Teachers Think about AI. Interviews*, Available from <https://www.aiedutainment.ai> (cited Aug 15 2026)
21. Mishra, P., *The Nostalgia Machine: Why Ed Tech Research Keeps Missing the Point*. Jan. 5 2026, citing Sparrow, B. et al., *Google Effects on Memory*. Science, 2011. Available from <https://punyamishra.com>. Sparrow B. et al., *Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips*. Science. Available from <https://www.science.org/doi/10.1126/science.1207745> (cited Aug 15 2026)
22. Aoun, J.E., *Robot-Proof: Higher Education in the Age of Artificial Intelligence*. MIT Press, 2017 (updated edition 2024); Northeastern University, *Robot-*

*Proof: President Aoun Outlines Plan for Reinventing Higher Ed in the AI Age.*  
Sept. 13 2017

23. Oakley, B. et al., *The Memory Paradox: Why Our Brains Need Knowledge in an Age of AI*. In: Rangeley, M. & Fairfax, N. (eds.), *The Artificial Intelligence Revolution*, 2025. Preprint available from <https://arxiv.org/abs/2506.11015>; SSRN 5250447 (cited Aug 15 2026)
24. Lee, V et al., *Cheating in the Age of Generative AI: A High School Survey Study of Cheating Behaviors Before and After the Release of ChatGPT*. *Computers and Education: Artificial Intelligence*, 7, 100253, 2024. Available from <https://doi.org/10.1016/j.caeai.2024.100253>; Spector, C., *What Do AI Chatbots Really Mean for Students and Cheating?* Stanford Graduate School of Education. Available from <https://ed.stanford.edu/news/what-do-ai-chatbots-really-mean-students-and-cheating>; International Center for Academic Integrity survey data (2002–2015, n>70,000: 64% admitted test cheating), as reported in The 74, *High School Cheating Increase from ChatGPT? Research Finds Not So Much*. Feb. 2024. Available from <https://www.the74million.org/article/high-school-cheating-increase-from-chatgpt-research-finds-not-so-much/> (all cited Aug 15 2026)
25. Nonis, S. et al., *An Examination of the Relationship Between Academic Dishonesty and Workplace Dishonesty: A Multicampus Investigation*. *Journal of Education for Business*, 77(2), 69–77, 2001 (n=1,051; moderate-to-strong correlation between academic and workplace dishonesty). Available from <https://www.tandfonline.com/doi/abs/10.1080/08832320109599052>; Guerrero-Dib et al., *Impact of Academic Integrity on Workplace Ethical Behaviour*. *International Journal for Educational Integrity*, 16(1), 2020 (n=1,203). Available from <https://link.springer.com/article/10.1007/s40979-020-0051-3>; Harding, T.S. et al., *Does Academic Dishonesty Relate to Unethical Behavior in Professional Practice?* Springer Nature 2004. Available from <https://link.springer.com/article/10.1007/s11948-004-0027>; Graves, S.M., *Student Cheating Habits: A Predictor of Workplace Deviance*. *Journal of Diversity Management*, 3(1), 2008. Available from [https://www.researchgate.net/publication/266243962\\_Student\\_Cheating\\_Habits\\_A\\_Predictor\\_Of\\_Workplace\\_Deviance](https://www.researchgate.net/publication/266243962_Student_Cheating_Habits_A_Predictor_Of_Workplace_Deviance) (all cited Aug 15 2026)
26. Sotiriadou, P et al., *The Role of Authentic Assessment to Preserve Academic Integrity and Promote Skill Development and Employability*. *Studies in Higher Education*, 45(11), 2132–2148, 2020. Available from <https://www.tandfonline.com/doi/full/10.1080/03075079.2019.1582015>; Nallaya, S. et al., *The Validity, Reliability, Academic Integrity and Integration of Oral Assessments in Higher Education: A Systematic Review*. *Issues in Educational Research*, 34, 2024. Available from <http://www.iier.org.au/iier34/nallaya.pdf> (both cited Aug 15 2026)

27. Zoting S., AI In Education Market Size, Share and Trends 2026 to 2035. Jan 5 2026. Precedence Research. Available from <https://www.precedenceresearch.com/ai-in-education-market>. Lower estimates: IMARC Group, AI in Education Market Report by Component, Available from <https://www.imarcgroup.com/ai-in-education-market> (\$79.6B by 2034) (cited Aug 15 2026)
28. Black, J., *How Microschools Became the Latest Tech Mogul Obsession*. WIRED. Aug. 18 2025, citing the Precedence Research market projection and Cowen, J., author of *The Privateers: How Billionaires Created a Culture War and Sold School Vouchers*. Available from <https://www.wired.com/story/silicon-valley-tech-moguls-microschools/> (cited Aug 15 2026)
29. Cowen, J., *The Privateers: How Billionaires Created a Culture War and Sold School Vouchers*, Harvard Education Press, 2024. ISBN -10 1682539105
30. Rasheed A., New UAE School Year Brings AI Curriculum, Unified Holidays, and Updated Exams. Gulf News, Aug. 24 2025, on Egypt's compulsory secondary AI subject and India's CBSE elective. Available from <https://gulfnews.com/uae/education/new-uae-school-year-brings-ai-curriculum-unified-holidays-and-updated-exams-1.500243448> ; Scales, J. et al., *Generation Ready: Scaling Safe, High-Quality AI In England's Schools*. Tony Blair Institute. Jan 22 2026. Available from <https://institute.global/insights/public-services/generation-ready-scaling-safe-high-quality-ai-englands-schools> <https://institute.global/insights/public-services/generation-ready-scaling-safe-high-quality-ai-englands-schools> (all cited Aug 15 2026)
31. Estonian Ministry of Education / AI Leap Foundation, Human Intelligence Alongside Artificial. Available from <https://tihupe.ee/en/>; Thian S., *Estonia Exercises Sovereignty in Its Ministry-Led AI Movement for Schools*. GovInsider, Apr. 9 2026. Available from <https://govinsider.asia/intl-en/article/estonia-exercises-sovereignty-in-its-moe-led-ai-movement-for-schools>; Regulations.ai, Estonia – National AI Education Initiative. *AI Leap — National AI in Education*. Available from <https://regulations.ai/regulations/RAI-EE-NA-ALNAEXX-2025>; on the GDPR firewall for student data: AP-sourced coverage, Feb. 2025. Available from [https://gigazine.net/gsc\\_news/en/20250227-estonia-launch-ai-leap-high-school/](https://gigazine.net/gsc_news/en/20250227-estonia-launch-ai-leap-high-school/) (all cited Aug 15 2026)
32. Ministry of Education, Singapore, EdTech Masterplan, Artificial Intelligence in Education. Available from <https://www.moe.gov.sg/education-in-sg/educational-technology-journey/edtech-masterplan/artificial-intelligence-in-education>; GovTech Singapore, *Inside Singapore's Digital Classroom. How AI is Supporting Teachers and Students*. Dec. 9 2025. Available from

<https://www.tech.gov.sg/technews/inside-singapore-digital-classroom-how-ai-is-supporting-teachers-and-students/> (both cited Aug 15 2026)

33. Center on Reinventing Public Education, *Shockwaves and Innovations: How Nations Worldwide Are Approaching AI in Education*. ASU Mary Lou Fulton Teachers College, Arizona State University. Aug. 2023. Available from <https://crpe.org/shockwaves-and-innovations-how-nations-worldwide-are-dealing-with-ai-in-education/> (cited Aug 15 2026)
34. Kwon J., *South Korea's AI Textbooks Fail to Make the Grade in South Korea After Rushed Rollout*. Oct. 15 2025. Available from <https://restofworld.org/2025/south-korea-ai-textbook/>; World Education Blog (UNESCO GEM), *AI Textbooks to Arrive in Korea — the Good, the Bad, and the Ugly*. Jan. 3 2025. Available from <https://world-education-blog.org/2025/01/03/ai-textbooks-to-arrive-in-korea-the-good-the-bad-and-the-ugly/>; Ministry of Education, Republic of Korea, *Briefing on the Plan for AI Digital Textbooks*. HIT 66864. June 9 2023. Available from <https://english.moe.go.kr/boardCnts/viewRenewal.do?boardID=254&boardSeq=95291&lev=0&m=0202&opType=N&page=2&s=english> (both cited Aug 15 2026)
35. Baron, N., *Who Wrote This? How AI and the Lure of Efficiency Threaten Human Writing*. Stanford University Press, 2023; Baron, N.S., *Why Human Writing Is Worth Defending in the Age of ChatGPT*. Literary Hub. Sept. 12 2023. Available from <https://lithub.com/why-human-writing-is-worth-defending-in-the-age-of-chatgpt/>; on the attribution history of the saying (O'Connor 1948, documented; Wallas 1926 and Forster 1927 as earliest instances; Faulkner versions secondhand, from 1989): Quote Investigator. Available from <https://quoteinvestigator.com/2023/01/26/think-read/>; King, S., *The Colorado Kid*. Hard Case Crime, 2005, p. 184. See as well AI Snacks with Romy&Roby, Episode 33, Ink Meet Code: *AI in Writing with Naomi S Baron*, transcript available on request (all cited Aug 15 2026)
36. Doudna, J. A., and Sternberg, S. H., 2017. *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*. Mariner Books.
37. Deutsch, D., 1997. *The Fabric of Reality: The Science of Parallel Universes and Its Implications*. Viking Adult.
38. Deutsch, D., 2012. *The Beginning of Infinity: Explanations That Transform the World*. Penguin Books.
39. AI Snacks with Romy&Roby, Episode 34, *Parenting in Code: The Snorble Story of Child-Centric AI with Mike Rizkalla*. Mar. 2025; Episode 59, *Kiddie Chatbots, Clone Employees, Digital You: Language AI with Rob LoCascio*, Jan. 2026 (all cited Aug 15 2026)

### Section 3: AI Literacy & Mid-Life Career — From Age Bias to AI Advantage

The cohort aged forty-five and older does not resist AI - it has discovered the tools without the training, and used them anyway. The real problem is a category error: the experience of this cohort is being priced as a liability when it should be priced as scarce capital. Here are *seven themes* describing the state of AI literacy and the experience of professionals aged 45 and older:

*First, the dominant discourse treats the 45+ cohort as resistant to AI, and the primary research proves the opposite.* In the edutAlnment survey of experienced professionals (n=1,065 with 221 full answers, February–April 2026), 93% have used Claude, ChatGPT or Perplexity; 52% use them daily or several times a day; and 91% agree that AI will fundamentally change how careers are built within the next five years.<sup>25</sup> Adoption is not the problem of this population.

*Second, the barriers this cohort faces are structural - and the constraint sits in organizations, not individuals.* Only 7% cite lack of confidence with the technology as a barrier to deeper use; cost and time dominate, and 44% report that their employer provides no AI upskilling support of any kind.<sup>25</sup> Microsoft's 2026 Work Trend Index finds that organizational factors account for roughly twice the AI impact of individual mindset; its "Blocked Agency" segment names workers with strong AI capability stuck inside organizations that have not caught up.<sup>26</sup> The dividend is not being captured because the architectures around this cohort are blocking it.

*Third, the AI job-loss and human replacement narrative is manufactured.* The AI platform CEOs who built the "jobocalypse" narrative (Altman, Amodei, Huang and Musk among them) are now quietly walking it back as the public backlash builds and the reversal data comes in.<sup>4</sup> The list of documented cases where roughly a third of AI-attributed layoffs have triggered subsequent rehiring, is growing. Roughly a third of those firms report that rehiring cost more than the original savings delivered.<sup>10</sup> The rehire scramble has begun, and the replacement clock is already running against the organizations that mispriced experience the first time; the manufactured scale of the panic is not. Job losses in some sectors are real and will continue. AI also redistributes work with many new jobs being created in services around AI.

*Fourth, the productivity and macro-employment data confirm the reversal.* The Return on AI Institute finds AI-attributed job cuts running at roughly 30 times the rate of measurable AI productivity gains,<sup>7</sup> and MIT's NANDA initiative reports that 95% of enterprise GenAI deployments produce no measurable ROI.<sup>8</sup> Stanford's ADP analysis shows employment for older workers in AI-exposed occupations has held steady or grown 6–9% since ChatGPT launched, even as younger workers in the same occupations declined 13–16%.<sup>13</sup> *Displacement pressure at the 40s-and-50s end of the workforce is not market-driven - it is a management choice,* and Carnegie's counter-narrative research reaches the same conclusion.<sup>13</sup>

*Fifth, the category error - mispricing experience - carries a \$9 trillion price tag. eGain and Deloitte (22 June 2026) project \$6.9-9.6 trillion in lost output from the Baby Boomer knowledge exodus over the coming decade; 85% of C-suite leaders name the exodus as a moderate to mission-critical threat - while 92% of organizations are failing to capture the knowledge before it walks out the door.<sup>14</sup>*

*Sixth, AI has increased demand for the very skills the 45+ cohort has spent decades developing. MIT Sloan's EPOCH framework, adapted by PwC's 2026 Global AI Jobs Barometer, names five ability clusters most complementary to AI: empathy, presence, judgment, creativity and vision. New tasks added to AI-exposed roles are 2.5 times more likely to require these than the tasks AI is taking over.<sup>6, 24</sup> These capabilities take decades to build and cannot be hired back at speed. The organizations losing them now will not find equivalents in the market when they go looking.*

*Seventh, organizations succeeding at the AI transition are building elastic architectures at scale.<sup>28</sup> Four organizations - FM, Cisco, Siemens and Unilever - are showing how to price experience correctly. They sit inside the 8% of enterprises not failing at AI deployment, and what they do differently is documentable, teachable and reproducible.*

*The argument throughout is that deep experience, correctly valued, is the scarcest capital in the AI transition. Those organizations that recognize it, will pull ahead.*

## **Mispricing Experience: The 45+ Cohort No Economy Can Afford to Lose**

The mispricing of experience is part of a wider pattern of human devaluation at civilizational scale, accelerated in the age of AI. Indy Johar, co-founder of Dark Matter Labs, nails it: *"The crisis is not human value. The crisis is human valuation."*<sup>47</sup> This section traces that pattern where it bites hardest - the 45+ cohort - and the five forces driving their devaluation on the balance sheet

## **The Demographic Lever No One Is Pulling**

Across the 1.4 billion people of the OECD's 38 member countries, and beyond them in China, Russia and now India, birth rates needed to maintain a steady population have fallen below the replacement level of 2.1 children per woman<sup>1</sup>, let alone sustain the growth that enterprises and economies rely on. No government anywhere has reversed the trend, and migration is the only near-term lever left to grow a working-age population. But the political appetite for it exists nowhere either.

So here is the scenario everyone frets about: a shrinking number of workers will be supporting a growing number of economically inactive people, and GDP per capita will slide by 2060 on the OECD's own numbers.<sup>1</sup>

But it does not have to be this way. We habitually frame an aging population as a problem - which only entrenches the mindset. Yet this is the healthiest, most

capable older generation in history. The most elegant fix to a labor shortage is hiding in plain sight: *design for the workers you already have, not the young ones you are waiting for.*

For developed-market economies, the average working life needs to be extended by 15% - an extra seven and a half years, raising the effective retirement age from 65 to 72½ on the US benchmark.<sup>2</sup>

*The longevity dividend is a gift to employers and treasuries alike, but few are pulling it. Unless the 45+ cohort is actively retained, retrained and engaged, the labor math won't work. Not for organizations staffing their AI transitions. Not for governments funding their welfare states. Not for economies chasing growth.*

Will organizations and governments act before the numbers overtake them?

### **The AI Transition: Future of Work ≠ Future of Jobs**

Depending on which agenda they are pushing, both the delusional-AI-enthusiasm camp and the apocalyptic-AI-anxiety camp are distorting predictions about AI's impact on business, society and employment. The scare tactics are driving public perception of AI-driven job loss well ahead of the data. Surveys find people estimate the rate of AI job destruction at roughly twice the measured number, and three times the rate among those personally affected.<sup>3</sup>

The AI platform CEOs crafted the “jobocalypse” narrative (Sam Altman (OpenAI), Jensen Huang (NVIDIA), Dario Amodei (Anthropic), and Elon Musk (xAI, owner of X and Grok) among them) to drive sales, valuations and political leverage.<sup>4</sup> Amid growing public backlash, data-center dread, technological sovereignty concerns, impending IPO calendars, a changing political atmosphere and cheaper competition from emerging entrants outside the US, they are now walking it back as the reversal data comes in.<sup>4</sup> Nobel-winning economist Paul Krugman named the mood shift plainly in June 2026: “*Why does everyone hate AI?*”<sup>5</sup>

Underneath the rhetorical climbdown sits an undeniable reality. *AI does not destroy work. It intensifies it and redistributes it.* Some jobs may disappear, but *tasks grow*, especially where AI is used to create new value, not just extract efficiencies. *The answer is to redesign work around the tasks that humans perform best alongside AI.*

PwC's 2026 Global AI Jobs Barometer, drawn from over a billion job advertisements across six continents, confirms the pattern. AI is producing a three-track global labor market: *professionalized jobs* reshaped to demand more expertise, judgment and leadership (radiologists, recruiters, air-traffic controllers); *democratized jobs* reshaped to demand less (software developers, loan officers, finance managers); and *low-exposure jobs* largely untouched (chefs, mechanics, construction workers) - a category growing around the AI sector. Jobs requiring specific AI skills are growing almost eight times faster than the wider jobs market; the AI-skills wage premium has reached 62%. AI-exposed entry-level roles are seven times more likely than non-

exposed roles to require oversight from traditionally senior skills, such as judgment and leadership.<sup>6</sup>

The long-term productivity case is meanwhile still failing to show up in the data. The Return on AI Institute finds AI-attributed job cuts running at roughly 30 times the rate of measurable AI productivity gains,<sup>7</sup> and MIT's NANDA initiative reports 95% of enterprise GenAI deployments are not yet producing measurable ROI.<sup>8</sup>

Challenger, Gray & Christmas has tracked approximately 120,000 US AI-attributed cuts since May 2023, with the monthly share accelerating from roughly 5% of total announced cuts in early 2024 to 26% by April 2026.<sup>9</sup> Demographic disclosure is not mandatory in all geographies, but the available evidence consistently points to employees in their 40s and 50s facing the hardest displacement when cuts do occur - age bias compounds AI-driven restructuring, and the same cohort struggles most to retrain or find equivalent roles.

A measurable share of these cuts are beginning to reverse. Research from Orgvue, CareerMinds and others documents the pattern: roughly a third of AI-attributed layoffs have triggered subsequent rehiring, and around a third of those rehiring organizations report the cost of restaffing exceeded the savings the layoff was meant to deliver.<sup>10</sup>

Several widely-reported reversals prove the pattern. Here are just two examples. Klarna declared in 2024 that its AI assistant was doing the work of 700 customer-service agents; by 2025, with customer satisfaction collapsing, the company was rehiring humans and routing support tickets to engineers and designers to plug the quality gap. CEO Sebastian Siemiatkowski conceded: "cost unfortunately seems to have been a too predominant evaluation factor... what you end up having is lower quality."<sup>11</sup>

Commonwealth Bank of Australia went further and admitted the mistake in writing: in July 2025, CBA declared 45 customer-service roles redundant on the basis that its "Hey CommBank" voice-AI bot would absorb the work. Within weeks the bank withdrew the redundancies, acknowledging that its initial assessment *"did not adequately consider all relevant business considerations and this error meant the roles were not redundant."*<sup>12</sup>

*The lesson is not that AI failed. It is where it failed.* Both reversals happened at the customer interface - the layer of work that requires creative interpretation, real-time problem-solving, judgment, empathy, and the relationship skills built across years of lived experience. AI can absorb the routine; it cannot read the room. The organizations cutting experienced workers from these functions are assuming they can rehire the capability later. But the capability is not in an LLM. It lives in the people.

Stanford's ADP-anchored analysis of the 2022–2025 employment cohort found that workers aged 22–25 in the most AI-exposed occupations experienced a 13–16% relative decline in employment since ChatGPT launched. The same analysis found

that employment for older workers in those same AI-exposed occupations held steady or grew between 6 and 9%. The Carnegie Endowment's April 2026 review reaches the same conclusion: "employment for non-AI-exposed professions and older workers has not changed since the release of ChatGPT in late 2022."<sup>13</sup> The displacement pressure on the 45+ workforce is not a market force; it is a management choice.

### **The price tag of the category error**

Deloitte Insights, in a joint report with eGain published June 2026, projects that the mass retirement of Baby Boomers - over 30 million Americans turning 65 in the next four years - will trigger the largest single transfer of institutional knowledge in business history. They put the economic value in lost output at *\$6.9 to \$9.6 trillion*.<sup>14</sup>

The report's main finding defines the crisis: 92% of organizations are failing to capture that knowledge before it walks out the door. The deep organizational memory of the engineer who knows why a system was built the way it was, or the account leader who understands the history behind a client relationship, will not naturally reaccumulate in the next generation where average tenure runs at around half that of Baby Boomers.<sup>14</sup>

That is the category error in hard numbers. And the firms making it will be racing to undo it in a rehire scramble within a decade. Three converging forces underpin this prediction: the institutional capability they have shed will be needed back; the demographic squeeze will leave fewer skilled workers in the market; the competition for that shrinking pool will price the rehire at a multiple of what the original cut saved.

### **The Cultural Bias the Science No Longer Supports**

Cultural bias against older workers is rarely articulated. It rarely needs to be. It is hardwired into hiring algorithms, promotion decisions, training budgets and severance calculations - the daily mechanics of who gets developed, who gets the next role, and who gets the chop first. The belief that perpetuates it is that digital technologies, and now AI-relevance, correlate with youthful adaptiveness and agility, and that older workers are too slow, too set in their ways, or too expensive to retrain given the tenure left before their "retirement date."

On cognition, today's older adults consistently outperform earlier-born older adults on measured ability, and age-related deficits are shrinking with each successive cohort, driven by lifelong gains in education, nutrition and healthcare. The IMF's microdata confirms it: *a 70-year-old in 2022 has the cognitive ability of a 53-year-old in 2000*.<sup>15</sup>

On AI tool use, the evidence points the same way. The randomized controlled trial behind GitHub Copilot's productivity claims found that developers with less programming experience, older programmers, and those who code more hours per day benefited most from the tool.<sup>16</sup> Older workers are among the cohorts who get the most out of AI, not the ones who struggle with it.

*The retreat from workforce training for older workers is now visible across every major developed economy. Across the OECD, the share of older workers receiving job-related training runs 30–40% lower than for prime-age workers - a structural under-investment that compounds with every passing year of working life. EU continuing-training spend per worker has been broadly flat in real terms since 2015, with persistent gaps for over-50s. US employer-provided training spend per employee fell approximately 12% in real terms between 2019 and 2024. The UK Skills England Annual Report 2026 documents the same pattern most precisely: British firms now spend half as much per employee on vocational training as the European average, total training expenditure in England fell 9% in real terms between 2019 and 2024, and per-employee spend is down 13% since 2019 across nearly every industry.*

A 2024 OECD study adds two further dimensions of the same pattern: *women carry higher structural exposure* because they are over-represented in the administrative, secretarial and customer-service roles most susceptible to automation; and *workers without tertiary education face exclusion from AI-augmented roles altogether* - missing the upside while still facing the displacement pressure.<sup>17</sup>

The strategic consequence is the one McKinsey Global Institute named in its May 2026 European study. Three-quarters of the skills European employers say they now need are used in both automatable and non-automatable work. These are human skills, applied alongside AI, not replaced by it. Demand for AI fluency has risen fivefold since 2023, and 75% of that demand sits in three occupation groups - computer & mathematical, management, and business & financial - the very groups the 45+ cohort dominates. *"AI is not replacing domain expertise,"* the authors concluded, *"but changing how it is applied."*<sup>18</sup>

This is the category error in operational form. The cohort with the deepest domain expertise is being priced as a liability, and starved of investment, at precisely the moment their expertise becomes the scarce resource AI needs to work alongside.

### **The 100-Year Life Has Arrived. The Societal Infrastructure Hasn't.**

AI Literacy is not only about individual capability - it is about the systems that determine whether capability can be deployed. When AI-driven displacement removes experienced workers from the labor market at scale, the cost cascades through every societal system designed for a different working life. Naming those systems matters because each is a leverage point for policy response.

When experienced workers are pushed out of the workforce in their 50s and early 60s, the cost cascades through housing, consumer credit, pensions, caregiving systems and the public purse. These systems were built for a working life that ended around 65 and a retirement that lasted around ten to fifteen years. *That world doesn't exist anymore. Citizens are now expected to support careers that are flattening or ending earlier than intended, and retirements that may last forty years.*

The chain reaction begins well before formal retirement. By age 50, salaries have plateaued, and restructuring and “efficiency” layoffs are hitting hard while the house still needs to be paid off, draining life savings or eating into pensions. Property accumulates but liquid reserves dry up. By the time the household reaches retirement, it owns property but has little flexibility. Property does not protect them. It immobilizes them.

*The financial system is out of step with the 100-year life and in urgent need of reimagining.* Globally, 65 has become the symbolic threshold where mortgage refinancing, borrower insurance and many financial products simultaneously expire, refuse new applicants, or become structurally unaffordable. The same banks that decline an €80,000 mortgage to a 72-year-old homeowner will happily extend high-interest revolving consumer credit to the same household - a pattern the French broker Arrago has documented at scale, with clients carrying €800 in monthly consumer-loan repayments against €1,200 in monthly income, on middle and lower-income wages.<sup>19</sup> *The age of 65 is not a biologically-grounded threshold but a political fossil: the calcified residue of a century of institutional systems too interlocked to redesign.*

When an insurer prices a client as “risky” at 65, they are perpetuating a convention rooted in the demographic reality of 1889 Germany. The Global Aging Institute's 2026 study “The Case for Lifetime Income,” sponsored by Prudential and drawing on data from five developed economies, arrives at the systemic verdict: *the systems were designed for people to die younger, and they have not caught up.* Pooling retirement income properly, the study finds, could deliver the same security at around 20% lower cost.<sup>20</sup>

### **The eldercare load is exiting the workforce alongside the AI layoffs.**

Family eldercare in the US has expanded to 63 million Americans providing care to aging adults - a nearly 50% increase since 2015, with 29% of them sandwich-generation caregivers supporting both children and aging parents simultaneously. 51% of sandwich-generation working mothers have left a job because of caregiving, and 49% report that caregiving has impacted their ability to advance in their career.<sup>21</sup> The cohort hit hardest is mid-career employees in leadership and specialist roles - precisely the people the AI conversation is also pushing toward the exit. *Many in the 45+ cohort are not leaving because they ran out of capability or ambition. They are leaving because the workplace infrastructure gave them no other path.*

### **The small-business succession cliff is a demographic crisis of its own.**

The European Commission issued a new formal “Recommendation on SME Business Transfers: on 23 June 2026, framing the crisis explicitly as a “*pressing demographic challenge*.”<sup>22</sup> The numbers don't give any room for a laissez faire attitude: 400,000 European businesses face ownership transitions every year, with millions more expected this decade as the first generation of post-war entrepreneurial founders reaches retirement age. In Korea, the average SME founder is now 62 years old and only 30% have any formal succession plan. In the

US, without a transition plan, an estimated six million businesses could close, taking twelve million jobs with them - and small businesses generate more than a third of all US business revenue and employ nearly half the workforce. In rural and rapidly aging areas of rural Europe, the UK and Australia, where small businesses can account for half of local employment, the risk is particularly acute. When ownership transitions fail, supply chains weaken, main streets hollow out, and communities lose essential services they cannot easily replace. Over 70% of family businesses globally fail to survive the transition to the second generation.<sup>22</sup> Germany, Switzerland, the Nordics and Korea are partly responding. The rest of the OECD footprint is in trouble.

The societal infrastructure problem is universal. The 70-year-life systems - pensions, employment contracts, benefit structures, mortgage products, caregiving frameworks, succession pathways - are creaking under a 100-year-life workforce.

### **The Skills AI Made More Valuable, Not Less**

The AI transition is dramatically increasing demand for the very skills the 45+ cohort has spent decades honing, while making it harder for the very same people to hold onto their jobs.

PwC's analysis of which human capabilities are most resistant to AI automation and most complementary to it adapts MIT Sloan's EPOCH framework. Five sets of abilities come through consistently:

- *Empathy and Emotional Intelligence* - compassion, understanding, building meaningful connections;
- *Presence, Networking and Connectedness* - physical presence, moving through social contexts, building trust;
- *Opinion, Judgment and Ethics* -working through ambiguous, open-ended or ethical scenarios;
- *Creativity and Imagination* - original thinking, improvisation, generating novel ideas;
- *Hope, Vision and Leadership* - grit, perseverance, initiative, strategic direction.

New tasks added to the most AI-exposed roles are 2.5 times more likely to require these abilities than the tasks AI is taking over.<sup>23</sup> Development Dimensions International, which has tracked leadership capability for over fifty years, names the same cluster in different words. The Chartered Management Institute's top three employability skills - team-working, critical thinking and problem-solving, and communication - sit inside it.

These are not abilities people are born with. The human brain learns through experience built over decades, through encountering enough situations to know which patterns matter, making enough mistakes to develop judgment, and building enough relationships to know how trust actually works. They are precisely the skills

that the 45+ cohort spent thirty years acquiring and hold in their highest, most-developed form. The window to transfer them through mentoring, redesigned work, and pairing AI Literacy with AI Fluency, is narrowing rapidly.

### **What the 45+ Cohort Is Actually Telling Us**

A respondent in our primary research, aged 40-to-50 and recently laid off in an AI-driven downsizing, wrote this in the open-text field:

*“If you're over 45 and you lost your job, it's not your fault. The system changed and is no longer rewarding you for the things it did before. ... If you can leverage 45+ managerial-level delegation skills to teach them to orchestrate teams of agents, this could really amplify them.”<sup>24</sup>*

This experienced employee has read the transition more clearly than most of the consultants and futurists shaping the debate. They have named the structural cause without self-blame, and articulated the recovery strategy in a single sentence: leverage decades of delegation experience to orchestrate teams of AI agents.

This respondent is one of 218 mid-career professionals aged 45 and over surveyed between February and April 2026 through edutAlnment and the AI Literacy Project's research circle.<sup>24</sup> Not a probability sample - *it captures the engaged subset*. Almost half are managers, directors or executives. Over a third have more than thirty years of professional experience. Three-quarters are based in Europe; one-fifth in North America.

The data reads as high engagement, not resistance. 93% have used Claude, ChatGPT or Perplexity. 52% use AI tools daily. 91% agree AI will fundamentally change how people find jobs and build careers within five years. 88% feel neutral, somewhat positive or very positive about AI and their careers.

What concerns them is not their own capability. It is the system around them. *“AI is filtering out CVs before they ever reach a human,”* one respondent wrote. The barriers to learning more are not psychological: only 7% cite lack of confidence with the technology; cost and time dominate. The cohort is paying for AI subscriptions out of pocket, learning on weekends, against the demands of the day job. 44% say their employer does not provide AI upskilling support.<sup>24</sup>

A second concern threads through their responses and foreshadows a deeper issue: a cohort using AI daily recognizes that tool fluency is not the same as judgment - the thing they have honed over many years - and worries about cognitive atrophy from over-reliance on the tools, and a human cognitive bandwidth that cannot scale infinitely.

Two 2026 enterprise studies independently confirm the diagnosis. Microsoft identifies a segment it calls Blocked Agency: workers with strong individual AI capability stuck in organizations that have not caught up.<sup>25</sup> WRITER's April 2026

enterprise survey of 1,200 senior executives and 1,200 employees reaches the same diagnosis from inside the C-suite, with one number that is a warning to be taken seriously by all leaders: *29% of employees admit to sabotaging their organization's AI rollouts when leaders impose mandates without strategy.*<sup>26</sup>

The constraint sits in organizational design - in culture, manager behavior, talent practices, how jobs are designed, and whether learning is built into the work or treated as someone else's problem. The eGain/Deloitte 92% gap is this same constraint at scale. *Pity for this cohort fixes nothing. Reconfiguring work fixes it.*

That reconfiguration has a name. The organizations adapting fastest are building what Anastassia Lauterbach's "Leading through Disruption" blog refers to as elastic organizations: *"Elastic organizations redesign their operating model to allow humans and agents to collaborate and deliver business results, not just their software stack."*<sup>27</sup> They blend permanent employees, contingent talent and AI agents as a single system that scales capacity up and down on demand, deployed in time-bound, intergenerational missions in which experienced professionals orchestrate, evaluate and mentor less experienced colleagues, with judgment flowing in both directions.

The elastic model has weaknesses. Net headcount savings are modest because oversight and exception-handling roles grow and require more expensive, senior people - the real value is throughput, not headcount elimination.

The binding constraint is cognitive: agents multiply faster than humans can manage them, with the failure pattern showing up as delegation creep, anchoring drift, and the fluency illusion the cohort itself flagged. Addy Osmani, Director at Google Cloud AI, calls this the Orchestration Tax: *"The price you pay for forgetting that while starting more agents is easy, your cognitive bandwidth doesn't parallelise."*<sup>28</sup> Deloitte's November 2025 "Unlocking Exponential Value with AI Agent Orchestration" report quantifies the risk at enterprise scale: more than 40% of today's agentic AI projects could be canceled by 2027 due to unanticipated cost, scaling complexity, or unexpected risks.<sup>29</sup>

With their proposed *"Octopus Organization"* of intelligence at the edges, coherence at the core,<sup>30</sup> Le-Brun and Werner offer a risk-management architecture: *distributing decision rights so that human oversight is not scaled beyond human capacity.*

What follows are four case studies of organizations already building elastic architectures of this kind, while integrating senior experience as part of their AI transformation.

#### **Four Organizations Pricing the 45+ Experience Correctly**

##### ***FM: Transfer Architecture — Treating Knowledge Transfer as Cognitive Capital Investment***

FM is a 190-year-old American mutual insurance company of around 6,000 people globally, many in highly specialized roles. Since 2021, over half of FM's Boomer

employees have left the organization to retire, so CHRO Robin Benoit responded to the strategic threat of losing invaluable institutional capability by treating knowledge transfer and succession as a prioritized capital investment rather than a side project. She launched the company's Accelerated Knowledge Advancement program (AKA) and with consultation, pulled fifty of its most experienced experts entirely out of their day jobs to become full-time mentors to ensure effective knowledge transfer to younger staff.

Benoit calls what she sees in the wider market “*AI washing*,” companies using AI hype as cover for layoffs while removing the hands-on experiences that layer and grow experienced workers in the first place.<sup>31</sup> FM is doing the opposite - *pricing experience correctly by investing in retaining it and amplifying its future growth*.

### ***Cisco: Tailored Learning Architecture — Senior Leaders Strained for Time to Learn***

Mid-career managers carry the broadest day-to-day operational load and are the most time-poor to learn AI by messing around and experimenting - a fact borne out by the 2026 Ipsos Generations Report on the experience of 3.6 million employees aged 16 to 66+ globally. For the 46–55 cohort, feeling under constant strain at work is their defining workplace challenge and the question on which they score lowest and lag most significantly behind every other age band.<sup>32</sup> This is also the same cohort identified earlier as the sandwich generation - juggling care for aging parents outside work hours, while carrying the heaviest day-to-day operational load inside them.

Over the past 18 months, tech giant Cisco has invested in AI tools and learning experiences designed to help all employees enhance their work and deliver measurable business results. To understand the impact of these investments, Cisco's People Intelligence team examined how employees engage with AI tools, drawing on surveys, interviews, focus groups, and data analysis. A stand-out insight from the research published in January 2026 was how employees learn AI makes all the difference. Director-level leaders at Cisco scored below mid-level employees on both confidence and satisfaction with the company's internal AI tool and skills training. Cisco's own analysts read it as a tailored-support problem: senior leaders need different learning opportunities to build the experience and confidence required to steer AI adoption. *A leader who is not immersed in and comfortable with the technology and fundamental design principles cannot credibly guide its use and advise its people.*<sup>33</sup>

### ***Siemens: Agent + Human Oversight Architecture — The Emerging Agentic Organization***

In a September 2025 analysis of early adopters unlocking significant value from AI transformation, McKinsey described the way in which the emerging “*Agentic Organization*” is architected.<sup>34</sup> In a culture characterized by continuous change and learning, talent is organized as:

- *M-shaped supervisors* - broad generalists fluent in AI, orchestrating agents and the hybrid workforce;
- *T-shaped experts* who hold deep specialist knowledge and safeguard quality;
- *AI-augmented frontline workers* whose tools embed senior judgment at the point of work.

This is the architectural response to the Orchestration Tax: distributing oversight across human roles so no individual is asked to scale beyond human cognitive capacity.

At Siemens, a multinational company focused on industry, infrastructure, transportation and healthcare, humans work alongside virtual and physical AI agents to create value serving billions of people. Siemens spends \$500 million a year reskilling its 320,000 people, with a 40-hour-a-year learning target built into the ESG (Environmental, Social, and Governance) framework. Chief People and Sustainability Officer Judith Wiese describes the philosophy: “*We redesign roles around the tasks that AI changes, before AI changes them unilaterally.*”<sup>35</sup> At their electronics factory in Erlangen, Germany, the “industrial copilot” is a workflow tool that embeds senior engineers’ troubleshooting patterns into a real-time AI assistant. This tool is enabling maintenance technicians to do work that previously required engineers, and younger and night-shift colleagues operate with the equivalent of a senior engineer at their elbow, on demand. Experienced engineers’ judgment flows down through the AI, not around it. This is an example of the EPOCH capabilities effectively transferred operationally, rather than theoretically.

For Siemens USA, the demographic urgency is equally explicit: nearly four million US manufacturing jobs need to be filled in the next five to seven years as Boomers exit, and to rapidly scale skills training at Fort Worth in Texas, Siemens hired schoolteachers and principals to skill their workforce while the factory was being built.

A second important insight from the Siemens transformation is that it is renegotiating the contract with its people. Global outplacement firm LHH’s 2026 research finds employability anxiety has now overtaken job security as the dominant concern across all ages, and this is also reflected in the sentiments expressed by our own survey respondents.<sup>36</sup> Siemens’s employer & employee co-determination model that puts works councils (German trade union representation) at the supervisory-board table has *shifted the labor conversation from “job security” to “employability” through continuous investment in future-relevant skills.*

### ***Unilever: Judgment-Layer Architecture - Built for Human Oversight***

Unilever’s global brands touch 3.7 billion users daily, and to protect their brands and users, they too have embedded a judgment layer into the redesign of AI workflows. At Unilever, AI doesn’t eliminate jobs; it reshapes them. Every AI deployment passes through a three-stage assurance process before deployment; screening for bias, transparency, and unintended harm. Making that call and drawing the line on

decisions that touch lives enough to require human judgment, is itself a judgment task. It is the kind of call that senior workers with years of pattern recognition experience make better than more junior ones, because they have lived long enough and have seen what unintended consequences look like before.

By mid-2024, 150 processes had been through it, and as much as half of them required adjustments. Operationalized at scale, this is what the design of good judgment infrastructure and good governance looks like — and it is built by people who have done the work long enough to know which edges matter.<sup>37</sup>

*The insight:* This is the operational antidote to the cultural bias documented earlier: experience is not being priced as a liability here; it is being deployed as the scarce resource AI deployment actually depends on.

Beyond AI governance, Unilever is also a leader in flexible employment design. Its U-Work program, launched in 2019, is a hybrid employment model that sits between traditional permanent employment and gig work. Participants remain Unilever employees, receiving a monthly retainer and core benefits — pension contributions, healthcare, sick pay - whether or not they are actively on assignment. When working, they earn additional assignment-based pay. The model is open to current employees and recent alumni and is designed to meet diverse life needs: phased retirement, caregiving, study, entrepreneurship, or working for other companies simultaneously. Now active in over 10 countries with more than 1,000 participants globally, U-Work provides the 45+ cohort options for gradual transition to retirement and in return, Unilever retains access to trusted, experienced talent while reducing reliance on external hiring and preserving institutional knowledge - the human tier of the elastic architecture, contingent talent built from people the organization already knows.<sup>37</sup>

### ***The UK Picture: A Wider Practice Emerging***

Beyond Unilever, a growing number of UK companies are offering greater flexibility for older workers. Just Group, a financial services provider, offers more part-time roles to attract mature staff members better placed to sell pensions. Cornerstone, a telecommunications firm, has introduced aged-care support for employees with caregiving responsibilities for elderly family members. John Lewis is encouraging more of its older employees to build skills by taking on senior apprenticeships. The UK is no longer an outlier on age-inclusive practice — it is emerging as a leadership cluster, with government, professional bodies and named employers pulling in the same direction, and leading the global conversation on longevity and the 100-year life from a societal point of view.<sup>37</sup>

## **From Diagnosis to Action: What Organizations, Policy-makers and the 45+ Cohort Can Do Now**

### ***For CEOs, CHROs and Boards: Six Organizational Actions***

#### ***1. Stop pricing experience as a liability.<sup>14</sup>***

*Levers:* Attribute an institutional-memory-loss risk value to every name on an AI-attributed redundancy list; conduct an age-bias audit of the last 24 months of AI-attributed layoffs; Board sign-off required on all cuts touching roles with rehire-reversal precedent — customer interface, complex judgment, senior stakeholder relationships.

*2. Treat knowledge capture as cognitive capital infrastructure — funded from operating budget, not innovation budget.<sup>14</sup>*

*Levers:* Apply the eGain/Deloitte five-step framework - authoritative knowledge foundation; structured expert capture; AI-enabled retrieval; governance; continuous renewal — embedded in operating budget; pair every senior expert nearing exit with a documented knowledge-transfer protocol before the transition, not after; build diagnostic infrastructure that determines which work AI can genuinely absorb before cutting the people who do it.

*3. Design returnships, senior apprenticeships, and AI-paired reskilling pathways at scale.*

*Levers:* partner with vocational and higher education institutions to build job-guaranteed, portable-credentialed pathways for the 45+ cohort; reference Meta's America's Workforce Academy is the blueprint;<sup>38</sup> run a 12-month AI-paired reskilling program for existing mid-career employees before external hiring; make senior apprenticeships<sup>37</sup> and returnships a standard; measure success on retention and internal mobility, not training-hours-delivered.

*4. Build the elastic organization deliberately, Orchestration Tax priced in.<sup>28</sup>*

*Levers:* Tie an internal price per intelligence unit <sup>39,40</sup> to business outcomes, not vendor-supplied token counts; Octopus architecture with intelligence at the edges, coherence at the core; distribute decision rights so no single human is asked to manage more agents than cognitive bandwidth allows.<sup>28, 30</sup>

*5. Redesign learning architecture for the time-poor senior leader and partner across sectors to guarantee quality.*

*Levers:* Buy the quality, don't build it; ring-fence learning budget per senior employee, redeemable across accredited providers and tracked for impact not compliance; sponsor industry-quality-assured Bootcamps rather than in-house tick-box modules; measure completion on demonstrated workplace application, not seat time.

*6. Hard-wire senior judgment into the AI deployment loop — and resolve the agent-ownership question before it litigates itself.<sup>42</sup>*

*Levers:* Treat Cognitive Sovereignty as a Board-level asset class alongside data governance and IP; embed senior judgment as the governance layer, not the

afterthought; update employment contracts, IP policies and AI governance frameworks to assign ownership of agent-encoded expertise before the expert exits.

### ***For Policymakers, Regulators and Governments: Five Policy Moves***

#### ***1. Update employment law for the AI hiring era.***<sup>42, 43</sup>

*Levers:* Extend bias-audit and transparency obligations to vendors as well as employers; enact statutory rights to meaningful human review of AI-driven hiring, promotion and termination decisions; demand regulator-accessible audit trails on training data and model outputs used in employment decisions.

#### ***2. Reboot financial products and benefit infrastructure for the 100-year life***

*Levers:* Decouple underwriting from age in pension, mortgage and insurance products; design portability into pensions, healthcare, training credits, parental and eldercare leave so entitlements travel across employment, fractional work and self-employment; alongside employers, pressure pension, insurance and banking partners to redesign products for the work of the future, not the past.

#### ***3. Build national strategic capacity for SME succession. Treat ownership transfer as critical demographic infrastructure.***<sup>22</sup>

*Levers:* Establish national-level funding for employee ownership trusts and founder transition financing; embed transfer-readiness diagnostics in SME support programs; treat SME succession as strategic infrastructure alongside energy, transport and digital.

#### ***4. Embed flexibility into the default employment relationship, and address the disproportionate burden women have carried for decades.***<sup>21,32</sup>

*Levers:* Offer tax incentives for employers offering phased retirement; enact statutory rights to fractional work and hybrid employment status (the U-Work pattern); introduce paid eldercare leave on par with parental leave and pension top-up rights for documented caregiving years; incentivize age-inclusive employers with reputational accreditation.

#### ***5. Treat AI Literacy as civic infrastructure. Fund it, embed it as foundational lifelong learning, design it to preserve critical thinking.***

*Levers:* Fund lifelong AI Literacy through public broadcasting, libraries, community colleges, adult-education frameworks and statutory school curricula; curricula covering AI system architecture, failure modes, manipulation patterns in conversational interfaces, and the cognitive habits that protect judgment; embed metacognition and critical thinking explicitly, not assumed; see the DOL February 2026 AI Literacy Framework as a starting reference.<sup>44</sup>

## ***For the 45+ Cohort Themselves: Four Personal Actions***

### ***1. Build AI Literacy first, AI Fluency on top.***

*Levers:* Treat AI Literacy as a life skill, not a training module; layer Fluency on top of a sound AI Literacy foundation; be wary of blindly investing in Fluency products that assume Literacy the buyer does not yet have.

### ***2. Become Octopus-like. Expand the mindset from title-role-job to orchestrator of a fluid skills bundle.***

*Levers:* Audit personal skills regularly against the EPOCH framework; build at least one revenue stream outside primary employment; treat professional networks as lifetime infrastructure, invested in before it is needed.

### ***3. Protect judgment actively.***

*Levers:* Use AI to extend judgment, not to outsource it or as substitute;<sup>45</sup> perform some routine tasks deliberately without AI to keep the underlying skills active; interrogate and cross-check AI outputs before acting, at all times.

### ***4. Practice through low-stakes experiments, and treat the transition as identity work as well as skills work.***

*Levers:* Enroll in free AI learning academies if budget is tight; learn-by-doing on inexpensive subscriptions; practice building personal agents to solve low-risk problems like weekly household menus, holiday research, shopping list optimization before deploying agents at work; nurture the identity work<sup>45</sup> of becoming AI-skilled: *someone who orchestrates AI, not someone replaced by it.*

## **In Closing: The Choice**

Lyndsey Simpson, CEO of 55/Redefined and author of *The Age Rebellion*, frames the choice in commercial terms: “The old story is that people hit their peak mid-career, then they start to plateau, and then they go into decline, and companies might quietly nudge them out, or intentionally encourage them to exit. But that story is now commercially dangerous, because midlife workforces are propping up all of the cracks in the backbones of our operations. They are the people with capability, judgment, institutional memory, customer relationships, understanding and empathy.”<sup>46</sup>

The 45+ cohort holds the EPOCH capabilities AI has made most valuable, at the moment AI most needs them. The replacement clock is running, the rehire scramble will grow, and meanwhile \$9 trillion of institutional knowledge is walking out the door daily. Experience cannot be manufactured. It can only be learned from the people who built it — and only while they are still here to teach it. *There is no third option.*

Two questions every CEO, business leader and policymaker must answer:

1. *What economic and human future am I enacting today with the decisions made about retaining, retraining or retrenching the experienced workforce?*
2. *And what is the business, societal and fiscal cost of that decision one year, five years, ten years from now?*

### Sources for Chapter 3\_Section 3

1. OECD Family Database, indicator SF2.1 (Fertility rates), 2024 data. Available from <https://www.oecd.org/en/data/datasets/oecd-family-database.html>; Eurostat Family and Fertility Statistics, February 2026 update; US CDC NCHS, *Births in the United States 2024*, Data Brief 535, July 2025; Australian Bureau of Statistics, *Births Australia 2024*, 15 October 2025. (cited Sept 18 2026)
2. Goldman Sachs (2025), demographic and workforce-extension analysis: 15% increase in working life, +7½ years, retirement age extension to 72½ on US benchmark. Available from <https://www.goldmansachs.com/insights/articles/aging-population-not-a-risk-to-the-global-economy-after-all> (cited Sept 18 2026)
3. Dilmegani C., AI Multiple research on the AI-driven-job-loss perception gap. Available from <https://aimultiple.com/fr/ai-job-loss>; Reuters / TechXplore, *AI chiefs walk back job apocalypse predictions*, 28 May 2026. Statements from Sam Altman (OpenAI), Jensen Huang (NVIDIA), Dario Amodei (Anthropic). Available from <https://techxplore.com/news/2026-05-ai-chiefs-job-apocalypse.html>; Musk E., statements on AI and future employment, Dubai World Government Summit (2023), xAI launch statements (2023), and subsequent public commentary. Coverage synthesized in TechCrunch archive: *AI's biggest names on jobs*. (all cited Sept 18 2026)
4. Reuters / TechXplore, *AI chiefs walk back job apocalypse predictions*, 28 May 2026. Available from <https://techxplore.com/news/2026-05-ai-chiefs-job-apocalypse.html>. Spencer, M, (August 2026) AI Supremacy Substack, *The AI Backlash is Accelerating*. Available from <https://open.substack.com/pub/aisupremacy/p/the-ai-backlash-is-accelerating-datacenters-2026> (cited Sept 18 2026)
5. Krugman P., *Why Does Everyone Hate AI?*, Substack, June 2026. Available from <https://paulkrugman.substack.com/p/why-does-everyone-hate-ai> (cited Aug 20 2026)
6. PwC, *Global AI Jobs Barometer 2026*. Analysis of over one billion job advertisements across six continents. Available from <https://www.pwc.com/gx/en/news-room/press-releases/2026/pwc-2026-ai-jobs-barometer.html> (cited Aug 20 2026).
7. Return on AI Institute (May 2026), report on AI-attributed job cuts vs measurable AI productivity gains. n=1,006 executives, 11 countries. Available from <https://roaiinstitute.com/white-paper/economic-maturity-for-artificial-intelligence/>
8. Challapally A., Pease R., Raskar R. & Chari P. (July 2025), *The GenAI Divide: State of AI in Business 2025*, MIT NANDA initiative. Available from <https://www.searchyour.ai/en/genai-divide-state-ai-business-2025-mit-nanda>
9. Challenger, Gray & Christmas, monthly job-cuts reports, May 2023 through April 2026. Available from <https://www.challengergray.com> (cited Sept 18 2026)
10. Orgvue (March 2026), workforce-planning research finding ~32% of organizations that made AI-attributed layoffs subsequently rehired; reported via Careermetrics, 30 March 2026. Available from <https://careermetrics.co.uk/blog/ai-layoffs-backfire-firms-rehiring-career-lessons-2026/>; Career minds (February 2026), survey of 600 HR professionals

- on post-layoff skills loss, capability gaps, and rehiring costs. Available from <https://careerinds.com/blog/cost-of-ai-layoffs>
11. Siemiatkowski S., CEO Klarna, interview on Bloomberg (2025). Klarna 2024 claim and 2025 reversal. Available from <https://www.bloomberg.com/news/audio/2025-09-10/bloomberg-talks-klarna-ceo-sebastian-siemiatkowski-podcast> (cited Sept 18 2026)
  12. Commonwealth Bank of Australia redundancy reversal (21 August 2025), 45 Customer Service Direct roles. Reported in *Information Age* (Australian Computer Society). Available from <https://ia.acs.org.au/article/2025/cba-reverses-ai-driven-job-cuts--admits--error-.html> (cited Sept 18 2026)
  13. Stanford ADP-anchored analysis of 2022–2025 employment cohort. Workers aged 22–25 in most AI-exposed occupations: –13 to –16% relative employment decline. Workers aged 40+ in same occupations: held steady or grew 6–9%. Available from <https://hai.stanford.edu/ai-index/2025-ai-index-report>; Carnegie Endowment for International Peace (April 2026), *The AI Labor Debate: Three Views on the Future of Work*, citing Stanford ADP-anchored pre-print (November 2025). Available from <https://carnegieendowment.org/research/2026/04/the-ai-labor-debate-three-views-on-the-future-of-work> (cited Sept 18 2026)
  14. Siegel E. (eGain) & Cahana E. (Deloitte), *The \$9 Trillion Knowledge Exodus: How Organizations Can Turn Baby Boomer Retirements Into Competitive Advantage*, Deloitte Insights, 22 June 2026. Available from <https://www.deloitte.com/us/en/insights/topics/talent/knowledge-management-plan.html> (cited Sept 18 2026)
  15. International Monetary Fund, *World Economic Outlook*, April 2025. Microdata on cognitive ability trajectories. <https://www.imf.org/en/publications/weo/weo-database/2025/april> (cited Sept 18 2026)
  16. Peng S. et al. (2023), *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot*, arXiv:2302.06590. Available from <https://arxiv.org/abs/2302.06590>; Skills England Annual Report 2026. Available from <https://www.gov.uk/government/publications/skills-england-annual-skills-report-and-sectoral-skills-needs-assessments-2026/skills-england-annual-skills-report-2026> (cited Sept 18 2026)
  17. OECD *Skills Outlook / Promoting an Age-Inclusive Workforce*, training-participation gap data. Available from [https://www.oecd.org/en/publications/promoting-an-age-inclusive-workforce\\_59752153-en.html](https://www.oecd.org/en/publications/promoting-an-age-inclusive-workforce_59752153-en.html); European Centre for the Development of Vocational Training (CEDEFOP) / Eurostat continuing vocational training survey data. Available from <https://ec.europa.eu/eurostat/web/microdata/collections-research/continuing-vocational-training-survey>; Association for Talent Development, *State of the Industry* annual reports, 2019–2024. Available from <https://www.td.org/product/research-report--2024-state-of-the-industry-talent-development-benchmarks-and-trends/192416>; OECD (October 2024), study on women's structural exposure to AI-driven displacement and education-level split. Available from <https://gwmh.org/oecd-policy-brief-algorithm-and-eve-how-ai-will-impact-women-at-work/> (all cited Sept 18 2026)
  18. McKinsey Global Institute (May 2026), European workforce / skills study. Available from <https://www.mckinsey.com/mgi/our-research/agents-robots-and-us-how-ai-reshapes-work-and-skills-in-europe> (cited Sept 18 2026)

19. Arrago (French reverse-mortgage broker), client data on senior credit patterns. Referenced via Longévité.xyz, *Le Patrimoine des seniors ne financera pas la dépendance*. Available from <https://www.longevite.xyz/p/le-patrimoine-des-seniors-ne-financera> (cited Sept 18 2026)
20. Global Aging Institute (2026), *The Case for Lifetime Income*. Sponsored by Prudential Financial. Data drawn from Australia, Japan, the Netherlands, the United Kingdom and the United States. Available from <https://www.globalaginginstitute.org/publications/reports.html> (cited Sept 18 2026)
21. AARP and National Alliance for Caregiving (July 2025), *Caregiving in the US 2025*. 63 million American adult caregivers; ~50% increase since 2015; 29% sandwich-generation. Available from <https://www.caregivingintheus.org/>; University of Phoenix Career Institute (2025), *Career Optimism Special Report: Moms in the Sandwich Generation*. 51% have left a job due to caregiving responsibilities. Available from <https://www.phoenix.edu/career-institute/sandwich-generation.html>; BioLife Plasma survey of 2,000 American caregivers aged 30–60 (2025). 49% report caregiving impacted career advancement. Available from <https://www.biolifeplasma.com/blog/sandwich-generation-costs-survey> (all cited Sept 18 2026)
22. European Commission (23 June 2026), *Commission Recommendation on Business Transfers of SMEs*. Available from [https://single-market-economy.ec.europa.eu/news/new-commission-guidance-will-make-it-easier-transfer-ownership-smes-2026-06-23\\_en](https://single-market-economy.ec.europa.eu/news/new-commission-guidance-will-make-it-easier-transfer-ownership-smes-2026-06-23_en) (cited Aug 20 2026); McKinsey Institute for Economic Mobility, *The Great Ownership Transfer: A New Era of Business Stewardship*. Available from <https://www.mckinsey.com/institute-for-economic-mobility/our-insights/the-great-ownership-transfer-a-new-era-of-business-stewardship>; Institute for Economic Democracy, *Europe Facing a Generational Transfer of Wealth: A Plan to Build Economic Autonomy and Ensure a Just Transition Through Employee Ownership*. Available from <https://ied.si/en/materials/europe-facing-a-generational-transfer-of-wealth-a-plan-to-build-economic-autonomy-and-ensure-a-just-transition-through-employee-ownership/> (all cited Sept 18 2026)
23. Loaiza I. & Rigobon R. (21 November 2024), *The EPOCH of AI: Human-Machine Complementarities at Work*, MIT Sloan School of Management working paper. SSRN abstract ID 5028371. Available from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5028371](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5028371). Framework profiled by MIT Sloan press (April 2026) and adapted by PwC in *Global AI Jobs Barometer 2026* (cited Sept 18 2026)
24. Anonymous respondent quote and edutAlnment / Cognoscenti / Savvitas / AI Literacy Project research circle, *AI Literacy Study: Experienced Professionals 45 Years Plus* (n=1,065, 221 full answers), primary research, February–April 2026. Survey available on request; Barriers to further AI learning: 71% cost, 50% time, 7% lack of confidence; UK adult AI confidence comparator: GOV.UK (28 January 2026), *AI Skills for Life and Work Summary Report*. Available from <https://www.gov.uk/government/publications/ai-skills-for-life-and-work-rapid-evidence-review/ai-skills-for-life-and-work-rapid-evidence-review> (all cited Sept 18 2026)
25. Microsoft (May 2026), *2026 Work Trend Index Annual Report*. Foreword by Dr Karim Lakhani, Harvard Business School. n=20,000 across 10 markets; Microsoft 365 telemetry; analysis subset n=16,971. Available from

- <https://www.microsoft.com/en-us/worklab/work-trend-index/agents-human-agency-and-the-opportunity-for-every-organization> (all cited Sept 18 2026)
26. WRITER × Workplace Intelligence (7 April 2026), *2026 AI Adoption in the Enterprise*. n=2,400 (1,200 C-suite + 1,200 employees), 5 markets, ~30 industries. Available from <https://writer.com/blog/enterprise-ai-adoption-survey/> (all cited Sept 18 2026)
  27. Lauterbach A. (12 March 2026), *The Elastic Organization*. Available from <https://www.linkedin.com/pulse/elastic-organization-dr-anastassia-lauterbach-5sbze/> (cited Sept 18 2026)
  28. Osmani A. (24 May 2026), *The Orchestration Tax*, personal blog; Osmani is Director, Google Cloud AI. Available from <https://addyosmani.com/blog/orchestration-tax/> (cited Sept 18 2026)
  29. Deloitte (18 November 2025), *Unlocking Exponential Value with AI Agent Orchestration*. Part of Deloitte Insights' Technology, Media and Telecom Predictions 2026. Available from <https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2026/ai-agent-orchestration.html> (cited Aug 20 2026).
  30. Le-Brun A.-S. & Werner F. (2025), *The Octopus Organization*. Available from <https://www.theoctopusorganization.com/> (cited Sept 18 2026)
  31. Morgan J. (5 May 2026), *Strengthening the Talent Pipeline for the Future of Work with the CHRO of FM*. Future of Work Leaders / LinkedIn Pulse. Interview with Robin Benoit, CHRO of FM. Available from <https://thefutureorganization.com/strengthening-talent-pipeline-future-work-fm-chro/> (cited Sept 18 2026)
  32. Ipsos (May 2026), *The Ipsos Generations Report 2026: Continuity vs Rupture*. 3.6 million employee responses, ages 16–66+. Available from <https://www.ipsos.com/en/ipsos-generations-report-2026-continuity-vs-rupture> (cited Sept 18 2026)
  33. Cisco Blogs (15 January 2026), *Cisco's Journey in AI Workforce Transformation*. Available from <https://blogs.cisco.com/news/ciscos-journey-in-ai-workforce-transformation> (cited Sept 18 2026)
  34. Sukharevsky A. et al. (September 2025), *The agentic organization: Contours of the next paradigm for the AI era*. McKinsey & Company. Available from <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-agentic-organization-contours-of-the-next-paradigm-for-the-ai-era> (cited Sept 18 2026)
  35. Wiese J. & Kerr W. R. (25 March 2026), *Siemens' Judith Wiese on engineering a global workforce transformation*. Managing the Future of Work podcast, Harvard Business School. Available from <https://www.hbs.edu/managing-the-future-of-work/podcast/siemens-judith-wiese-on-engineering-a-global-workforce-transformation> (cited Sept 18 2026)
  36. LHH (March 2026), *Employability Anxiety: The New Employability Crisis*. Available from <https://www.lhh.com/en-us/insights/the-new-employability-crisis> (cited Sept 18 2026)
  37. CIO.Inc (12 August 2025) *Unilever's AI Push: From Shop Floor to Culture Core - Part 2* Available from <https://www.cio.inc/unilevers-ai-push-from-shop-floor-to-culture-core-part-2-a-29183>; Telegraph.co.uk (14 May 2026) *The companies pulling out all the stops to hire over-50s*. Available from

- <https://www.telegraph.co.uk/money/jobs/career-advice/companies-hiring-midlifers/> (cited Sept 18 2026)
38. Meta (8 June 2026), *America's Workforce Academy*. Available from <https://about.fb.com/news/2026/06/americas-workforce-academy-free-skilled-trade-training/>. \$115M first-year investment, paid stipend, job guarantee, NCCER-portable certification; Meta Careers, *Return to Work* program. Available from <https://www.metacareers.com/careerprograms/pathways>; Industry returnships overview, May 2026, covering JPMorgan, Goldman, Microsoft Leap, Intuit. Available from <https://blog.theinterviewguys.com/returnship-programs-for-2026/> (cited Sept 18 2026)
  39. Palmer S. (7 June 2026), *Price Per Intelligence Unit*. Available from <https://shellypalmer.com/2026/06/price-per-intelligence-unit/> (cited Sept 18 2026)
  40. *Workforce Planning with AI Agents*, wfmlabs.org wiki. Available from [https://wiki.wfmlabs.org/wiki/Workforce\\_Planning\\_with\\_AI\\_Agents](https://wiki.wfmlabs.org/wiki/Workforce_Planning_with_AI_Agents) (cited Sept 18 2026)
  41. *Singapore SkillsFuture program. National credit account system*. Available from <https://www.skillsfuture.gov.sg> (cited Sept 18 2026)
  42. Cohen N. (28 April 2026), *Can You Take Your AI Agents With You When You Leave A Job?*, Forbes. Available from <https://www.forbes.com/sites/niritcohen/2026/04/28/can-you-take-your-ai-agents-with-you-when-you-leave-a-job/>; Flynn G. (2026), *The Special Sauce Problem: Who Really Owns Your AI Agents?*, LinkedIn Pulse. Available from <https://www.linkedin.com/pulse/special-sauce-problem-who-really-owns-your-ai-agents-gareth-flynn-aljvc/>; Brcic M., *The Memory Wars: AI Memory, Network Effects, and the Geopolitics of Cognitive Sovereignty*, SSRN abstract ID 5308691. Available from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5308691](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5308691); companion Research Square preprint. Available from <https://www.researchsquare.com/article/rs-9145237/v1> (cited Sept 18 2026)
  43. *Mobley v. Workday, federal collective certification*, opt-in deadline 7 March 2026. Age Discrimination in Employment Act claim; ACLU Colorado / EEOC settlement, March 2025, age thresholds in AI hiring algorithm; Colorado AI Act (SB 24-205), effective June 2026; NYC Local Law 144, automated employment decision tools bias audit requirement; *Over 45 and looking for a job? AI thinks you might be too old*, phys.org, 29 May 2026. Available from <https://phys.org/news/2026-05-job-ai.html> (cited Sept 18 2026)
  44. US Department of Labor (February 2026), *AI Literacy Framework*. ETA Training and Employment Notice TEN 07-25. Available from <https://www.dol.gov/agencies/eta/advisories/ten-07-25> and <https://beta.dol.gov/ai-ready>; \$243M apprenticeship funding announcement; AARP Research, *How AI is Impacting the Future of Work Among Adults Age 50-Plus*, August 2024, updated May 2026. Available from <https://www.aarp.org/pri/topics/work-finances-retirement/employers-workforce/workforce-trends-older-adults-artificial-intelligence/> (cited Sept 18 2026)
  45. McNamara S. et al. (2025), *AI-Induced Replacement Dysfunction (AIRD)*, proposed clinical construct; Delphi-validated thematic analysis of Indian IT professionals (2025), layered psychological disruption from AI-induced

displacement. Available from <https://www.cureus.com/articles/407877-artificial-intelligence-replacement-dysfunction-ai-a-call-to-action-for-mental-health-professionals-in-an-era-of-workforce-displacement#!> (cited Sept 18 2026)

46. Simpson L. (2025), *The Age Rebellion*. CEO of 55/Redefined. Available from <https://www.telegraph.co.uk/money/jobs/career-advice/companies-hiring-midlifers/> (cited Sept 18 2026)
47. Johar I. (30 August 2026), Being Human: Being Priced Out of the Future — AI, capital and the crisis of human valuation, Substack essay. Indy Johar is co-founder of Dark Matter Labs. Available from <https://indyjohar.substack.com/p/being-human-being-priced-out-of-the> (cited Sept 18 2026)

## Section 4: AI Literacy in Small and Midsized Businesses (SMEs)

The businesses carrying most of the world's employment and innovation - small and mid-sized enterprises - have adopted AI faster than almost anyone predicted and understood it more slowly than almost anyone admits.

SMEs are running AI-supported workflows, tolerating unsanctioned AI use by their employees, buying foundational models from three geopolitical blocs, and struggling to budget for climbing compliance costs - while telling us, in their own words, that what they most need is an answer to the most basic question of all: *what is AI, actually?*

The result is *a global SME economy being reshaped by a technology its users deploy daily but cannot yet explain, audit, or defend.*

Why does AI Literacy matter in the context of SMEs?

The complexity economist César Hidalgo in "The Infinite Alphabet" says: "*We tend to think of knowledge as something that accumulates passively, like books on a shelf. But productive knowledge, the kind that drives the economy, is more like a muscle. If you stop exercising it, it weakens.*"<sup>1</sup> The economy, in Hidalgo's framework, is at some level a giant game of Scrabble: each firm holds a set of letters - capabilities, skills, knowledge - and the products and services it can generate are the words it can form. AI hands out an enormous new supply of some letters while leaving others as scarce as ever. That's why businesses that understand AI and deploy it to increase their competitiveness expand their alphabet.

We identified several themes describing the state of AI in SMEs:

First, as in every other segment we researched, *AI usage is near-universal. AI Literacy is not.* We call it the *deployment paradox*. In the edutAlnment survey of SMEs (n=915),<sup>2</sup> only 4.7% of businesses don't use any AI, and 89.2% view AI as strategically vital over the next 3-5 years.<sup>2</sup> 58.1% say many of their employees use AI tools daily - yet only 26.5% actively train their staff on AI.<sup>2</sup> People know the product, not the mechanism and what it is based on. Their picture of AI is therefore narrow, and when the product changes - a new vendor, a new agent framework, a novel risk to address - the foundation is too thin to transfer.

Second, *SMEs cannot afford to be AI-illiterate.* They cannot absorb failed AI pilots, cannot outbid Big Tech for talent, and - as we will see - face compliance costs designed with far larger balance sheets in mind. They want AI to work directly to resolve competitive pressures: *automate routine tasks* (59.8%), *introduce AI as a substitute in certain tasks* (49.8% - a figure that will intensify the workforce design conversation), *train new hires faster* (43.0%), and *support decision-making in complex roles* (41.2%).<sup>2</sup> Across the survey, we see contradictory answers that point to a *lack of deep strategic abilities to decide where and how to deploy AI*. On the one hand, companies fear losing control over their data or falling victim to a cyber-attack;

on the other, the majority of SMEs tolerate the use of shadow AI. On the one hand, SMEs fear a knowledge exodus; on the other, they state that AI could substitute employees in many tasks. On the one hand, the skill gap is emphasized as a major concern; on the other, only about a quarter of companies actively train their workforce to be AI-ready.<sup>2</sup> What we witness is *strategic and operational confusion* - and confusion leaves little chance of competing successfully in increasingly complex markets. The practical answer, developed in the roadmap that closes this chapter, is that an SME cannot out-spend an enterprise on AI - but it can out-decide one.

Third, in Europe only the UK is well represented in our data. With the latest survey wave in September 2026, European-headquartered companies account for 32.1% of responses: 24.1% in the UK, 5.1% in the EU and 2.9% in other European countries.<sup>2</sup> UK professionals form the second-largest group of individual respondents (22.1%), behind the United States (39.9%) and ahead of Singapore (19.6%) and Saudi Arabia (18.5%). EU-headquartered companies, however, remain strikingly underrepresented. While the imbalance may partly reflect the survey's distribution channels, it is consistent with a well-documented pattern: *European SMEs engage with AI later and less intensively than their international competitors*. Roughly 34% of US companies use AI in some form, against an EU-wide average of about 20%, and at the individual level 43% of US workers report using AI at work compared with 26–36% across major European economies.<sup>3</sup> More than half of large European companies have yet to scale a single major AI investment,<sup>4</sup> and among European SMEs, non-use remains the norm: Eurostat counts AI use in only 17% of small EU enterprises, against 55% of large ones.<sup>5</sup> Whether measured at the level of firms, workers or investment, the reluctance we observe in our own response data mirrors a broader European hesitancy - one that economists increasingly link to the productivity gap that has opened between Europe and the United States since the mid-1990s.<sup>3, 6</sup> The UK respondents who joined the survey confirm the pattern from the inside: they are markedly more cautious than their peers in the US, Asia and the Middle East - a difference we examine in a dedicated section below.

Fourth, *beneath the AI technology adoption story lies a people crisis*. Asked about their biggest workforce challenges, edutAlnment survey respondents named *difficulty hiring skilled staff* (44.7%) and *the skills gap in digital, data, and AI competencies* (43.3%) most frequently, ahead of high turnover (39.9%).<sup>2</sup> Lack of junior talent troubled 37.5%, and the knowledge drain from an aging workforce concerned 32.6%; asked directly, 76.3% were very or somewhat concerned about experienced staff retiring without employees or apprentices to replace them. The generational anxiety was explicit: 33.3% were very concerned about new generations entering the workforce and 45.4% somewhat concerned, worrying most about skill gaps and professionalism (48.5%), work values and expectations (46.0%), and communication (45.7%).<sup>2</sup> AI accelerates the decades-old phenomenon of *automation bias* - *the systematic tendency to over-rely on automated systems and under-scrutinize their output, even when it is wrong*.<sup>7</sup> Our respondents sense the danger: 22.7% name the fear of losing skills or of overdependency on AI as a concern.<sup>2</sup> A junior employee whose first years are spent approving AI outputs rather than producing work will appear productive but learn little.<sup>8</sup> The design principle that follows is the same one we set out for classrooms: *use AI as an amplifier on top of internalized knowledge, never as a replacement for building it*. The honest use of AI in a small business is not

always an immediate relief on the cost side. Overall, automation bias is another variation of the cognitive debt risk we have already discussed across the whole AI Literacy Study.<sup>9</sup>

All these pressures frame *the four frontiers of the HR function - workforce design, recruiting, training, and culture* - developed later in this section of the AI Literacy Study.

Fifth, only 33.7% of surveyed SMEs see AI as a tool to *capture and codify expert knowledge* before experienced staff retire, despite identifying knowledge exodus as one of their top challenges.<sup>2</sup> We call this the *knowledge-capture blind spot: a gap between problem awareness and solution design*, and one of the clearest strategic openings the edutAlnment survey surfaced.

Sixth, *data protection, privacy, and security risks at 49.7% are the single most-cited concern* in the entire edutAlnment survey.<sup>2</sup> At the same time, 61.1% of companies *tolerate use of shadow AI*.<sup>2, 10</sup> The security dimension of this contradiction, and its low-cost remedies, is examined in the section on shadow AI and data protection.

Seventh, for SMEs, *AI implementation costs are not an entry barrier but a recurring shock*. High implementation costs rank among the most-cited obstacles to AI adoption in small firms,<sup>11</sup> and among adopters the majority miss their AI cost forecasts – roughly one in four by more than 50% - with unexpected costs already forcing spending freezes and cancelled initiatives,<sup>12</sup> a margin of error a mid-sized firm cannot absorb the way a large corporation can. The edutAlnment data sharpen this picture in an important way: the binding constraint our respondents describe is not the price of the tool but the cost, in scarce expertise and time, of making the tool work inside a real business – and precisely these costs are invisible at the moment of purchase. No license fee captures the hours spent redesigning a workflow, cleaning the data the tool depends on, or teaching a team to verify its output. Economists have long observed this pattern with general-purpose technologies: the visible technology investment is only the tip of a much larger, unmeasured investment in complementary skills, processes and organizational change, which is why returns lag adoption and why firms systematically underestimate what adoption truly costs.<sup>13</sup> The remedy, developed below, is to stop treating AI as a product one buys and start treating it as infrastructure one composes – increasingly from capable open-weight models.

Eighth, *the workforce must be trained for the AI era - and in Europe, this necessity has become law*. Across OECD surveys, roughly half of SMEs report that their employees lack the skills to use generative AI effectively - one of the foremost barriers to adoption, ahead of funding and infrastructure.<sup>14</sup> New official data confirm the gap: in the UK, limited AI skills, expertise or knowledge is the second most-cited barrier to adoption (60%),<sup>15</sup> and only 11% of firms with ten or more employees report that more than half of their workforce has received AI-related training.<sup>16</sup> Article 4 of the EU AI Act obliges every organization deploying AI in a professional context, regardless of size, to take measures on the AI literacy of its staff and of anyone operating AI on its behalf – applicable since February 2025, softened in wording but not in timing by the July 2026 Digital Omnibus, and supervised and enforced by

national market surveillance authorities from August 2026.<sup>17, 18</sup> What the law now mandates is what competitiveness has demanded all along; the obligation, its practical implications, and the difficulty of choosing a training partner in a noisy vendor market are discussed in the following pages.

Ninth, *SMEs can benefit disproportionately from local ecosystems - but ecosystems do not work by themselves; they require leadership*. This point was not delivered by the edutAlnment survey; it is clear from our implementation practice and the research. SMEs collaborating with universities and research centers have been found up to six times more likely to adopt AI,<sup>19</sup> and universities, schools, startups and public programs - from European Digital Innovation Hubs to national AI vouchers - stand ready as on-ramps. But on-ramps carry only the firms that drive onto them: someone must convene the local table, scope the pilot, sign the intern, and settle the IP in the contract. The SME whose leadership does so becomes the firm the region's talent, research and funding find first - a theme we develop in the roadmap that closes this chapter.

## **The AI Deployment Paradox in SMEs**

In 2026, edutAlnment surveyed business professionals in small and mid-sized companies across North America, Europe, Asia and the Middle East; 915 professionals opened the survey.<sup>2</sup> All respondents were employed full-time. 63.1% of their companies already use AI-supported workflow management, 56.4% use chatbots and virtual assistants, and 53.0% use proprietary large language models such as ChatGPT, Copilot, Gemini, or Claude; 22.1% already work with open large language models such as Llama, and 27.2% report digital avatars of their executives or best employees. Agentic AI - systems that coordinate autonomous multi-step tasks without constant human prompting - is already deployed by 23.8%, a remarkable figure for a technology that was barely production-ready 24 months ago. Only 4.7% use none of the above.<sup>2</sup>

And yet, when the same respondents were asked what everyone in their company should learn about AI, the most-requested topic - chosen by 59.1% - was: *What is AI, and what are the key technologies powering AI applications?*<sup>2</sup>

We call this the *deployment paradox*, the SME sibling of the *practitioner paradox* we identified among educators. One respondent described the dilemma plainly:

*"We really don't know what AI can do or how it could help us. A conversation with someone who knows what the various tools are currently could help with that. It's evolving so fast it feels pointless to commit to one."*<sup>2</sup>

Company size told a nuanced story: the largest single cohort of our survey (28.6%) came from companies with 500-1,000 employees, straddling the upper boundary of most SME definitions, while the genuinely small end (1-50 employees) represented 26.0%. Revenue was distributed remarkably evenly from under USD 1M (15.6%) to over USD 1bn (5.1%). Industry breadth was equally telling: Technology and Telecommunications led (23.2%), but Construction (16.2%) and Financial Services

(10.2%) followed, with sectors from Hospitality to Pharmaceuticals represented - confirming that AI literacy has become a cross-industry imperative. Demographically, respondents cluster in the 30–44 band (74.5%), with 20.4% aged 45–59.<sup>2</sup>

Our respondents came from businesses - construction firms, clinics, factories, marketing agencies - whose value was built on domain craft that predates the AI era, and they were genuinely interested in AI. In the wider economy, OECD data show that SMEs use AI increasingly. In 2024, only 11.9% of small firms (10–49 employees) and 20.4% of mid-sized firms used AI, against roughly 40% of large enterprises - an adoption gap that translates directly into a productivity and competitiveness gap.<sup>20</sup> In the OECD's 2026 D4SME survey of over 2,000 firms across twelve countries, 61% of SMEs were using at least one AI-enabled application - but 76% of those adopters remained *AI novices*, relying on simple tools for isolated uses rather than understanding or integrating the technology.<sup>21</sup> How adoption is measured is itself part of the literacy problem. In the United States, the Census Bureau's Business Trends and Outlook Survey put business AI use at 19.8% in May 2026, with no significant growth among firms with fewer than 20 employees;<sup>22</sup> the U.S. Chamber of Commerce, using a broad definition of AI tools, finds 66% of small businesses using AI;<sup>23</sup> and the Federal Reserve's Small Business Credit Survey finds 46% of employer firms using AI - but only 7% of these users have fully integrated it.<sup>24</sup> The headline numbers diverge with the definition; the depth problem does not.

Our respondents' companies were headquartered primarily in the US/Canada (37.8%) and the UK (24.1%), followed by “Other Asia” including Japan, Korea, India, and Southeast Asia (13.3%), and the Middle East & Africa (10.8%); the EU accounted for only 5.1% and other European countries for 2.9%.<sup>2</sup> Their reach is wider than their headquarters: 48.3% serve customers or work with partners in the US/Canada, 35.2% in the UK, 22.2% in the EU and 19.7% in other Asian markets. The four countries most represented among individual respondents - the United States (39.9%), the United Kingdom (22.1%), Singapore (19.6%), and Saudi Arabia (18.5%) - mirror the global corridors of AI investment and adoption ambition. *The near-absence of EU-headquartered firms, given Europe's regulatory push on AI, is itself a data point, and it repeats a finding from our wider study.* Continental European SMEs were strikingly hesitant to discuss their AI programs at all, compared with peers in the US, the UK, the Middle East, and Asia - a sign not only of hesitant leadership but of a regional deficiency in having any broad dialogue on “anything AI.”<sup>2</sup>

In the coming months, edutAlnment will roll out its SME survey to further geographies and industrial sectors in collaboration with data collectors.

Strategic urgency in the edutAlnment survey was unambiguous: 37.4% said AI is *critical* - they must adopt it to survive - and a further 51.7% called it important for competitive edge. Combined, 89.2% view AI as strategically vital over the next 3–5 years; only 3.1% said their business model does not require it. The greatest potential is perceived across multiple business functions: customer service (47.2%), product development and R&D (44.8%), cybersecurity and risk (42.7%), sales and marketing (41.3%), and data analysis (40.6%), with finance (38.5%) and operations (37.4%) close behind and HR (31.1%) further back.<sup>2</sup> Only 16.1% see staff cost reduction as a

major area of potential - a curious contrast with the 49.8% who want AI to substitute for employees in certain tasks.<sup>2</sup>

Asked about company's leadership, respondents describe an engaged executive layer: 60.4% said their leadership team actively drives AI strategy, and a further 27.9% said leadership understands the basics and supports pilots - nearly 90% leadership engagement, a result that demolishes the narrative of AI adoption being blocked by executive inertia.<sup>2</sup> 53.4% report a clear AI strategy tied to business outcomes with a 3-5-year outlook - yet 37.2% are still experimenting to figure out how AI can help, 24.8% implement AI for a few use cases without an overarching strategy, and only 18.8% have a dedicated function supporting AI implementations; 2.7% forbid AI altogether.<sup>2</sup> Because respondents could select several answers, many companies apparently claim a clear strategy and open-ended experimentation at the same time. AI that pre-dates the GenAI wave is in our data: machine learning appears in 30.2% of companies and predictive analytics in 26.5%.<sup>2</sup>

But the shape of that knowledge is thin where it matters. The same organizations whose leaders *actively drive AI strategy* told us their whole workforce still needs to learn what AI *is* (59.1%), what an AI agent is and how it works (43.8%), and how to use AI and agents to make work easier or faster (43.8%).<sup>2</sup>

## Europe and Other Regions: What Is Different?

What separates Europe - including the UK - from the other regions in our data is not *whether* SMEs use AI, but how deeply, how confidently and under which rules. The respondents added in the latest wave of the edutAlnment survey are almost entirely UK-based (61 of the 276 respondents to open questions, against none in the earlier wave), which allows an indicative comparison with the earlier, predominantly US, Singaporean and Saudi cohort.<sup>2</sup> Roughly one in five of the newly added respondents called AI *critical* to survival; about half said their leadership actively drives AI strategy and that their company tolerates shadow AI; only about three in ten described themselves as very familiar with AI regulation and equipped with a compliance plan (51% before); only around one in eight had deployed agentic AI, while about half feared job losses and employee resistance. Three-quarters deploy US models and the same one-third is using Chinese ones.

Governmental and institutional statistics describe the same breadth without depth: Eurostat counts AI use in 20% of EU enterprises - 17% of small, 30% of medium-sized and 55% of large firms;<sup>5</sup> the ONS puts UK adoption at 30% in September 2026, with 29% among firms of fewer than ten employees against 51% among those with 100 or more;<sup>25</sup> the US Census Bureau measured 19.8% in May 2026;<sup>22</sup> and Singapore's SME adoption jumped from 14.5% to 23.4% within a year.<sup>26</sup> The European Investment Bank finds EU firms on par with US firms in generative AI use (37% vs. 36%), but only 55% of EU adopters use AI in more than two business activities, against 81% in the US.<sup>27</sup>

What is genuinely distinctive about Europe is the rulebook: the EU is the only region in our survey where AI literacy is a legal duty of every deployer - since July 2026 an

obligation to support its development rather than to guarantee a result<sup>17, 18</sup> - while the UK still has no comprehensive AI law and relies on voluntary upskilling, with a government goal of training 10 million workers by 2030,<sup>28</sup> and six in ten US small businesses worry about a patchwork of state privacy and AI laws.<sup>23</sup> The conclusion for European SMEs is uncomfortable but useful: their hesitancy is less a technology gap than a gap in confidence and depth - precisely the gap AI literacy must address and close, if implemented as a strategic and operational exercise, not as a compliance box-checking.

## Shadow AI and Data Protection

Only 23.2% of edutAlnment respondents want everyone in their company to learn how to use AI ethically and responsibly, and only 23.1% name ethical and reputational risks as a concern.<sup>2</sup> Data protection and privacy, and security risks at 49.7% are the single most-cited concern in the entire edutAlnment survey, followed by fear of job losses and employee resistance (40.2%), lack of internal expertise (36.7%), regulatory compliance (35.0%), and the speed of the AI skill gap relative to innovation pace (29.4%); one in five (20.3%) consider AI overhyped.<sup>2</sup> We call this the *data protection prism*. AI reaches most SMEs filtered through data risk before any other consideration can arrive - similar to the *cheating prism* through which teachers first see AI.

61.1% of companies said their employees use personal AI applications, and the company simply tolerates it. Only 30.2% have strict rules, and 8.7% do not know.<sup>2</sup> We call this posture *tolerance by default*, and it is not a neutral choice. It is shadow AI: sanctioned work moving through unsanctioned models, with customer records, source code, and financial documents landing on servers the organization does not control. When we shared the tolerance figure with a fellow data scientist, his response was: "I would be shocked if it's that low." He is right - non-response bias means we cannot know what the silent companies would have said, and no policy stops an employee from picking up a phone, getting an answer, and pasting it into an e-mail.<sup>2</sup> Fresh UK evidence supports his intuition: in Deloitte's 2026 survey of 25,000 UK working adults, 31% of generative AI users said they use it without their employer's knowledge, and 17% pay for at least one tool out of their own pocket - an estimated £958 million a year - while around half of users have received no formal training.<sup>29</sup>

Levi's contribution to this AI Literacy Study<sup>30</sup> reframes what shadow AI actually is in security terms. His *tank versus drone* whitepaper argues that three decades of perimeter-first security - firewalls, VPNs, endpoint agents, ever-thicker stacks of controls - built organizations that resemble tanks: expensive, armored, and optimized for predictable threats, now facing swarms of cheap, adaptive, AI-enabled attacks that probe continuously at near-zero marginal cost.

For the SME, two of Levi's findings matter most. The first: AI is an amplifier, not an inventor. Across the government and standards-body research he synthesizes - the G7 Cyber Expert Group, ENISA, the World Economic Forum, NIST - most breaches still trace to familiar failings: unpatched systems, mismanaged credentials, misconfiguration, accumulated technical debt. AI simply exploits them faster, more

cheaply and without fatigue. Verizon's 2026 data makes the point obvious: exploitation of unpatched vulnerabilities overtook credential abuse as the leading initial access vector, present in 31% of breaches, even as the median time to patch known-exploited vulnerabilities rose from 32 to 43 days. The same report found shadow AI detections rose fourfold in a year, making unsanctioned AI the third most common non-malicious insider action in breach data.<sup>31</sup> Every “temporary” exception, orphaned account and unmanaged SaaS tool is a door, and shadow AI is the latest and fastest-multiplying species of that debt - with one property no earlier shadow IT ever had: data submitted to an external model cannot be reliably retrieved, deleted or audited afterwards.<sup>32</sup>

The second finding: the governance gap is the cheapest gap to close. IBM's breach research found that 97% of organizations that suffered an AI-related security incident lacked proper access controls on their AI systems, and 63% had no formal AI governance policy at all.<sup>33</sup> These are not million-euro deficiencies; they are decisions nobody made. Levi's conclusion is the security translation of this study's central claim: AI does not remove the need for basic discipline - it raises the cost of neglecting it, and it demands that the basics now operate at machine speed, because processes calibrated to human pace become liabilities once the adversary is software.<sup>30</sup>

### ***Recommendations for SMEs on Cybersecurity and Data Protection***

Levi's five shifts for enterprise security leaders translate to SME scale as six practical steps:

*First, know what you have - on one page.* An SME does not need an asset-management platform; it needs a living, one-page inventory of systems, admin accounts, SaaS subscriptions and AI tools in actual use, reviewed quarterly. Shadow AI enters this list by asking, not auditing: an amnesty-style survey - *what tools do you actually use?* - converts invisible risk into a governable inventory, and the finding that 45% of employees are now regular AI users on corporate devices guarantees the list will not be empty. When it comes to employees using shadow AI, the governance answer is not the hammer to forbid AI use. A wiser CIO looks for the people willing to admit they use shadow AI and asks how the organization has underserved them: can we create a safe, air-gapped sandbox where they experiment with anything they want, and then come back for help? The response to shadow AI is some combination of compliance, serving the underserved need, and channeling what will happen anyway into a place where it can be seen.

*Second, close the governance gap before buying anything.* A two-page AI usage policy - *which tools are sanctioned, what data may never enter a public model, who approves new tools, what to do after a mistake* - plus access controls on the AI tools already in use, addresses the two deficiencies present in 97% and 63% of AI-related incidents. This is the highest-leverage, lowest-cost step available; it costs a workshop, not a budget.

*Third, pay down the debt that drones exploit.* Patch on a calendar - the attacker's tempo is faster than the 43-day median, and an SME's small estate is its advantage, because there is less to patch. Enable multi-factor authentication everywhere, delete orphaned accounts on the day someone leaves, and cancel unused subscriptions. Levi's rule - *fix complexity before scaling AI* - is easier for an SME than for any enterprise, because the terrain is smaller.

*Fourth, buy machine-speed defense; do not build it.* The enterprise builds telemetry pipelines and automation playbooks; the SME's equivalent is *choosing vendors whose products already contain the defensive AI* - mainstream business platforms with built-in phishing detection, anomaly alerts and automated quarantine - and, where budgets allow, a *managed detection-and-response service*. The selection question to ask any vendor comes straight from Levi: show me how your AI detects and contains, not just alerts.

*Fifth, treat AI systems as systems to secure.* Any AI tool the firm deploys - chatbot, avatar, agent - gets the same basic scrutiny as any other critical system: who can access it, what data flows into it, what it is permitted to do, and what the vendor's answers are to the OWASP LLM Top 10 risk categories.<sup>34</sup> For SMEs this is a procurement checklist, not an engineering discipline - and it connects directly to this study's one-primary-vendor recommendation *against the frankensteinization of agentic AI*.

*Sixth - the step Levi lists last and this study puts first - train the people.* His re-skilling recommendation for security analysts becomes, at SME scale, an equivalent to the EU AI Act Article 4 baseline: everyone touching AI understands its fundamentals, limits and failure modes. In a firm with no security team, the workforce *is* the security team.

## **Implementing AI with Open vs. closed APIs**

The edutAlnment data show some SMEs are making genuinely strategic infrastructure decisions. 82.3% of respondents use or plan to use US-based foundational models (OpenAI, Anthropic, Google) - but 33.6% also use models incepted in China, citing stronger support for local regulations (38.2%), better performance in the Chinese language and local markets (35.7%) and strategic diversification (35.0%) well ahead of lower cost (26.9%); and 29.3% use EU-based or open-source models such as Mistral. Asked about national models in general - models built in their own country or region - respondents cite regulatory alignment with rules such as GDPR, the EU AI Act or HIPAA (52.3%) and local-language performance (49.5%).<sup>2</sup> Concerns about Chinese models center on data sovereignty (40.3%), geopolitical and sanctions risk (38.2%) and the lack of transparency in training data and model behavior (37.1%).<sup>2</sup>

The pattern matches what we observed globally and read in secondary literature.

The capability argument for a closed-API monoculture has collapsed: by 2026 the leading open-weight families - Llama, Mistral, Qwen, DeepSeek, Gemma, Phi, joined

in August 2025 by OpenAI's own gpt-oss, its first open-weight release since GPT-2 - compete directly with closed APIs. Mozilla's inaugural *State of Open Source AI* report measures the average capability gap at 3.3 points on Chatbot Arena, with open models at parity in coding, instruction-following and general knowledge, closed models retaining a lead only in advanced reasoning and complex agentic work - and open-weight models now routing a majority of production tokens on OpenRouter, the largest neutral model marketplace.<sup>35</sup> The geography of the shift is itself a finding: China and East Asia lead open-source AI adoption at 89%, and in February 2026 Alibaba's Qwen overtook Meta's Llama as the most-downloaded model family on Hugging Face - deliberate industrial policy, as Mozilla notes, with China's State Council directive explicitly promoting open-weight releases.<sup>36</sup> SMEs worldwide have begun implementing these models, adapted to local context with full data sovereignty; one of our own cases involved exactly such a model adjusted to the environment of a European mid-sized company. Lauterbach talks about the new *Cold War over Hot Tech* since 2019,<sup>37</sup> and today it means internationally active companies may need to serve both American and Chinese standards to compete in key markets. One counterweight keeps the picture honest, and it is Mozilla's own: capability parity has not become deployment parity. While 79% of developers use open models, only 53% of open-model teams reach production, against 63% for closed - a gap the report attributes not to model quality but to missing operational tooling, and enterprise buyers accordingly now prioritize licensing terms (31%) and data ownership (26%) over raw capability.<sup>35</sup> That gap between what is possible and what firms can operate is, precisely, a literacy gap - and the buyers' shift toward control over capability is the market pricing that literacy in.

We have already discussed practical recommendations on implementing open-weight models in SMEs in Chapter 2 of this AI Literacy Study.

## **AI Implementation Costs and Budgets**

The edutainment respondents are willing to invest in AI. The largest single AI budget bracket was USD 50K-200K annually (29.9%), followed by USD 200K-1M (25.6%) and USD 10K-50K (19.2%); 8.9% spend more than USD 1M, only 3.6% had no dedicated AI budget at all, and 71.9% said they have a methodology for estimating the ROI of AI projects.<sup>2</sup> Asked which cost components they expect, respondents put training and change management for staff first (46.6%), ahead of licensing and subscription fees (44.1%), data preparation and management (42.0%) and customization and integration (41.3%) - while only 13.9% budget for monitoring and auditing AI models and agents.<sup>2</sup>

With average SMB spending on AI tools estimated in the tens of thousands of dollars, nearly two-thirds (64.4%) of our respondents budgeting above USD 50K marks this sample as unusually committed adopters - a self-selection the survey design anticipated, and one consistent with the global pattern: BCG finds every industry planning to increase AI spending in 2026,<sup>38</sup> and NVIDIA reports 86% of enterprises raising AI budgets, with only 2% cutting.<sup>39</sup> More striking is a second figure: 71.9% of respondents say they possess a methodology for estimating the ROI of AI projects.<sup>2</sup> It stands in sharp contrast to the independent evidence: across 2025-26 surveys, only 45% of organizations can quantify AI ROI at all, a quarter of

initiatives met expected returns, only 7% of leaders in KPMG's research report established ROI, and 56% of CEOs in PwC's global survey report no significant financial benefit from AI to date.<sup>40</sup>

Two readings of edutAlnment data are possible, and both serve this study's argument - AI literacy is imperative. The generous one: our self-selected, budget-committed respondents genuinely are ahead of the market in measurement discipline. The skeptical one: "having a methodology" is a self-assessment, and what most firms call an ROI methodology measures productivity anecdotes and cost avoidance rather than margin or revenue - *fluency mimicking literacy*, in the vocabulary of this study. *The gap between the 71.9% who claim a methodology and the 11% who can forecast their AI spend within ten percent is, either way, a measurement literacy gap.*

More revealing is what respondents said would make AI more affordable.

Easier integration with existing systems topped the list (46.3%), followed by shared or collaborative AI infrastructure (41.6%), public funding and grants (41.3%), and pre-built industry-specific solutions (41.3%). Lower licensing costs came only fifth, at 38.4%.<sup>2</sup> We call this the *integration inversion*, and it is one of the edutAlnment *most actionable findings for platform providers and policymakers alike: the binding constraint is not the price of the tool but the cost, in scarce expertise and time, of making the tool work inside a real business.* This is, at bottom, *a knowledge problem that manifests as escalating cost - a pattern economists have documented for every general-purpose technology, where the visible technology investment is only the tip of a much larger, largely unmeasured investment in complementary skills, processes and organizational change.*<sup>13</sup> The OECD reaches the same conclusion from the other direction: across G7 surveys, roughly 50% of SMEs report that their employees lack the skills to use generative AI effectively - the single greatest barrier to adoption, ahead of funding and infrastructure.<sup>14</sup>

*The knowledge problem is aggravated by the unforgiving economics of building and deploying AI, which vendors prefer to keep hidden.* The AI market deliberately lacks price transparency that we have already discussed in Chapter 2. Pricing is fragmented across per-seat subscriptions, per-token API metering, hybrid base-plus-overage tiers and bespoke enterprise contracts – and the cheapest option at trial is frequently the most expensive at twelve- or thirty-six-month scale. The most damaging costs are usually invisible at the moment of purchase: retrieval-augmented workloads that send large document chunks with every call can generate several times the token volume of simple chat,<sup>41</sup> and organizations that fail to model context-window usage routinely underestimate their actual AI spend. The consequences are no longer tolerable. Sixty-two percent of organizations surveyed in 2026 by Mavvrik said an unexpected AI cost materially altered a business decision in the past year; among them, 40% required board-level escalation, 33% implemented emergency spending freezes and 25% delayed or cancelled an AI initiative.<sup>12</sup>

Large and mid-sized companies alike budget for the license and forget the meter, with no financial hedging of compute exposure of the kind a CFO would apply to FX or commodity risk. The first hedging instruments are only now emerging.<sup>42</sup>

Finally, the edutAlnment survey suggests SMEs are aware of AI regulations and are worried about additional costs. 45.9% said they are very familiar with AI-related regulation and have a compliance plan, and a further 40.1% are starting to assess its impact; only 5.4% have not engaged at all. Compliance spending is already material - 29.7% expect USD 50K-200K annually and 23.7% expect USD 200K-1M - and the top challenges are documenting AI systems (41.6%), understanding which rules apply (40.9%), and the high cost of external legal support (38.4%). What they want is telling: affordable legal and compliance advisors (44.4%), practical guidance tailored to SMEs (43.7%), templates for documentation and risk assessment (43.7%), and affordable cybersecurity technology (43.4%).<sup>2</sup>

## **Elastic Organizations and the Novel Challenge for the People Function**

Our survey shows that the pressure to redesign organizations around AI is already acute. Beneath the technology adoption story lies a people crisis: respondents named difficulty hiring skilled staff (44.7%) and the skills gap in digital, data and AI competencies (43.3%) as their biggest workforce challenges, ahead of turnover (39.9%), with lack of junior talent (37.5%) and the knowledge drain from an aging workforce (32.6%).<sup>2</sup> The generational anxiety was explicit: 33.3% were very concerned and 45.4% somewhat concerned about new generations entering the workforce, worrying most about skill gaps and professionalism (48.5%), work values and expectations (46.0%), and communication (45.7%).<sup>2</sup>

What respondents want AI to do maps directly onto these pressures: automate routine tasks (59.8%), introduce AI as a substitute in certain tasks (49.8% - a figure that will intensify the workforce conversation), train new hires faster (43.0%), and support decision-making in complex roles (41.2%). But only 33.7% see AI as a tool to capture and codify expert knowledge before experienced staff retire - even though 76.3% of the same respondents are concerned about retirements without successors.<sup>2</sup> We call this *the knowledge-capture blind spot*: a gap between problem awareness and solution imagination.

The edutAlnment survey was not designed to examine any single business function within SMEs. Yet across its responses - on skills gaps, role substitution, generational concerns and knowledge loss - one function keeps emerging as the addressee: HR. The people function now owes the business two things at once - an organization fit for AI, and a workforce that turns AI into competitive advantage - delivered across four frontiers: how the workforce is designed, how it is recruited, how it is trained, and how its culture adapts.

### ***Frontier 1: Strategic Workforce Planning and Design***

*HR has traditionally planned for a workforce made of people; it must now plan for a workforce in which humans work alongside AI agents and smart machines – a redefinition of the labor force itself, independent of company size.* Microsoft's 2025 Work Trend Index describes the emerging “frontier firm”, in which every employee becomes a manager of agents, and finds that a large majority of leaders expect agents to be integrated into their workforce strategy within eighteen months.<sup>43</sup> AI, in

this context, is becoming the infrastructure through which a new type of labor is integrated. Almost no HR function is equipped to shape it: only 5% of HR teams describe themselves as well-prepared to integrate AI at all,<sup>44</sup> and only 8% of HR leaders believe their managers have the skills to use AI effectively.<sup>45</sup>

For the future of work and of the workforce, we favor the term *elastic organization*: an operating model in which permanent employees, contingent talent and AI agents work together as a single flexible system that scales capacity up and down on demand – deployed in time-bound, intergenerational missions in which experienced professionals orchestrate, evaluate and mentor, with judgment anchored in humans and insights flowing in both directions. On paper, SMEs already live closer to the elastic model than large companies: flat hierarchies, generalists wearing multiple hats, an existing reliance on freelancers and partner networks, and decision paths short enough that a redesigned workflow can go live in weeks rather than quarters.

The elastic model, however, brings four major challenges:

*Net headcount savings are modest.* Oversight and exception-handling roles grow, and they require more expensive, senior people. The real value of elasticity is throughput and resilience, not headcount elimination - a correction to the 49.8% of our respondents planning AI as a substitute in certain tasks.<sup>2</sup>

*The binding constraint is cognitive.* Agents multiply faster than humans can manage them; the failure pattern shows up as delegation creep and the fluency illusion. Google engineering leader Addy Osmani calls it the Orchestration Tax: starting more agents is easy, but "your cognitive bandwidth doesn't parallelize" - and the tax is the price of forgetting it.<sup>46</sup> For an SME whose entire oversight capacity may be one founder and one operations lead, the tax bites earlier and harder than in any enterprise.

*Agentic AI quietly creates what our study calls the vanishing loop.* Human-out-of-the-loop becomes the default rather than the exception, and the workflow becomes effectively unauditable - the log line reads, in effect, "permission temporarily elevated to complete task," and no ticket, no manager's authorization, no reconstructable decision trail exists. Modern compliance regimes presume an identifiable responsible officer; an agent-run process offers regulators no such person.<sup>47</sup> An SME with a three-person finance team cannot afford to discover this during an audit.

*New digital assets raise new governance questions.* Who owns the avatar that encodes an expert's judgment when the expert leaves?<sup>48</sup> edutAlnment recommends answering this question in the contract, before any problem surfaces. We further recommend selecting one primary vendor for AI agents and avatars, to avoid what we call the frankensteinization of agentic AI - an estate stitched together from mismatched tools that no one can govern as a whole.<sup>49</sup>

The design answer to all four challenges is the same: distribute decision rights so that human oversight is never scaled beyond human capacity - intelligence at the edges, coherence at the core. This is the "octopus" model<sup>50</sup> of the individual and the

organization that Killian develops in her section on experienced professionals aged 45 and older, and elastic organizations will need precisely this design to succeed. SMEs may lack the budgets of the large firms in Killian's case studies, but they hold one strategic advantage their large peers cannot buy: speed.

AI becomes a driving competitive force when speed meets knowledge and experience. And yet Killian's chapter documents the category error of the decade: experience being priced as a liability when it should be priced as scarce capital. A Deloitte Insights report co-authored with eGain projects a \$6.9-9.6 trillion knowledge exodus as Baby Boomers retire; 85% of C-suite leaders call it a moderate-to-mission-critical threat - yet 92% of organisations fail to capture the knowledge before it leaves.<sup>51</sup> Stanford's analysis of ADP payroll records shows why the design matters: since late 2022, early-career workers in the most AI-exposed occupations have seen a 13-16% relative decline in employment, while experienced workers in the same fields held steady or grew - a pattern the authors interpret as AI substituting for codified knowledge, but not for the tacit knowledge accumulated over decades.<sup>52</sup>

Demographic trends amplify the knowledge exodus - and for SMEs specifically, the workforce equation has a third term the survey barely registered: succession. The European Commission's June 2026 guidance on SME business transfers names the issue explicitly as a "pressing demographic challenge" with the potential to affect millions of companies over the next decade;<sup>53</sup> in Korea, about one-third of SMEs are run by owners aged 60 or older, and 28.6% of these owners have no successor;<sup>54</sup> in the US, some six million small and mid-sized businesses face ownership transitions by 2035, with effective transitions estimated to keep up to twelve million jobs in place;<sup>55</sup> and by the oft-cited estimate, some 70% of family businesses worldwide do not survive the transition to the second generation.

Knowledge capture, junior development and succession are not three problems. They are one problem - and AI, deployed with strategic foresight, is the first technology that can address all three at once.

### *Case Study: AI Stepping in for a Retiring Workforce*

In 2025, edutAlnment was tasked with implementing AI at a family-owned European industrial hidden champion - an OEM for textile machines, selling globally - whose experienced factory workforce was retiring, whose apprenticeship program could no longer recruit, and whose experienced workers had never documented how they did what they did.

The project began not with an AI model but with the archaeology of a factory floor: video capture of workflows across 47 factory jobs, using off-the-shelf tools; conversion of the videos into structured workflow documentation; process redesign in workshops with the workers themselves; and AI-generated training documents and videos, so that what had lived only in experienced hands became teachable. Only then came the automation - process-mining and workflow platforms integrated with the ERP and specific OEM tools - and only then the workforce redesign: FTE projections, transition paths, and explicit decisions about knowledge capture,

including AI avatars of the firm's best people. This was not a technology project with an HR annex; it was a knowledge-preservation project in which AI happened to be the instrument.

## ***Frontier 2: Recruiting***

AI has entered recruiting from both sides of the market at once: employers deploy it to source, screen, rank and interview candidates, while candidates deploy it to write applications, polish CVs and rehearse interviews - sometimes carrying an AI-assisted persona deep into the hiring process. Two senior talent leaders interviewed for the AI Literacy Study sharpen the picture from the hiring side.

Nic Weatherhead, VP of People Operations and Talent Acquisition at The Stepstone Group, described AI already threaded through recruitment - sourcing tools that can carry inherent bias, note-taking tools that analyse candidates and interviewers alike, stack ranking of applicants - while candidates themselves increasingly deploy AI to apply for roles, sometimes getting far into the process. Her central warning was a literacy warning: organisations should not rely solely on procurement and legal teams to manage compliance, because assessing any tool requires a deep understanding of the workflow and of where the human belongs in the loop – and teams asked to do more with less, adopting AI on top of years of headcount reductions, risk burnout on the ground.<sup>56</sup>

A second executive, GM of Global Talent Acquisition for EMEA & Asia at a large technology group, was blunter still about the vendor landscape: many providers are “selling the dream” of AI tools without fully understanding or explaining the legal and regulatory environment their customers operate in - and SMEs in particular, with a different risk profile, often prioritize cost reduction over compliance or inclusive hiring. He expects a reduction in recruitment roles over the next five years - consistent with the 49.8% of edutAlnment respondents planning AI substitution in certain tasks - even as today's efficiency tools mostly offset a surge in AI-generated applications. On the hopeful side, he noted that well-trained tools can reduce bias relative to human decision-making, provided the data is good and the humans stay calibrated.<sup>56</sup>

Recruitment is, as Weatherhead put it, both an art and a science - process and data on one side; assessment, critical thinking and essential human skills on the other. That formulation generalizes to almost every SME function about to be transformed: the science can increasingly be delegated; the art cannot - and it is the art that literacy protects. The research names the art precisely: the EPOCH cluster - *empathy, presence, opinion/judgment/ethics, creativity, and hope/leadership* - comprises the capabilities most resistant to automation and most complementary to it; new tasks in AI-exposed roles are 2.5 times more likely to require them.<sup>57</sup>

## ***Frontier 3: Workforce Training***

Respondents across edutAlnment and other surveys genuinely want to learn about AI. Even if one argues that they represent only a minority of society as a whole -

survey samples self-select: not everyone who received the survey opened it, and not everyone who opened it completed it - their voices should not be ignored, precisely because a motivated minority is all a multiplier effect requires. Educating and training those willing to learn can spread competence through businesses far beyond the individuals trained.

Asked openly what would make their company more AI-literate and confident, our respondents overwhelmingly requested practical, role-specific, hands-on training built on real business use cases - and named the noise of the training market itself as an obstacle: "There are so many AI trainings out there and there's a challenge in figuring out what is right for our business at every level of need."<sup>2</sup>

Five requests recur from edutAlnment survey data:<sup>2</sup>

*Practical, role-specific, hands-on training* – "Short workshops showing how AI applies directly to daily tasks (not just theory)" - overwhelmingly the most common answer, echoed across responses in English, Arabic and German, with several respondents specifying micro-learning formats.

*Clear guidelines and governance* - simple rules on when and how AI can be used, including data privacy and security expectations, so that employees "stop improvising policy one prompt at a time."

*Leadership modelling* – "If leadership doesn't use it, nobody will," as one Arabic-language respondent put it, in an unusually complete diagnosis that concluded: companies don't fail at AI because of the technology – "they fail because of culture, fear, and unclear use."

*A trusted guide through the noise* - a coaching relationship with vendors "instead of 'just take it and implement it,'" industry-specific seminars ("specific to the construction industry," wrote one respondent), and *someone to help* "choose, use, monitor, audit the various AI tools."

*Psychological safety* - perhaps the most human finding: several respondents, notably in the Arabic responses, framed the goal of training as removing the *awe and fear* of AI - creating a culture of safe experimentation in which small mistakes are the tuition, not the crime.

Two things are notable about this list. First, it is cheap relative to the stakes: nothing on it requires a data science team. Second, it is nearly identical to what teachers asked for in our section on AI Literacy in Teachers - short, practical, role-relevant, guilt-free learning - which suggests that the format of effective AI literacy is converging across segments even as some content diverges.

In the European SME context, our story diverges sharply from what we see in the education sector.

Where teachers wait for governments to develop a vision, SMEs serving European customers already face one - in the form of law. Article 4 of the EU AI Act obliges providers and deployers of AI systems to take measures on the AI literacy of their staff and anyone operating AI on their behalf, proportionate to their roles and the context of use. The obligation has applied since 2 February 2025; the July 2026 Digital Omnibus recast it from “ensuring a sufficient level” of AI literacy to “supporting its development” - an obligation of effort rather than of result - but the duty itself and the enforcement calendar were untouched: national market surveillance authorities begin supervising and enforcing it from 2 August 2026.<sup>17, 18</sup> The definition the law relies on is, notably, a genuine literacy definition, not a training checkbox: skills, knowledge and understanding that allow people to make an informed deployment of AI systems and to gain awareness of opportunities, risks and possible harm.<sup>58</sup> Regulation itself, in other words, has begun mandating what this study argues for - with one caveat our wider study states bluntly: a legal obligation discharged with a one-hour compliance video is about mimicking fluency, not building knowledge and understanding - that is, literacy.<sup>10, 59</sup>

There are three practical implications for SMEs:

First, *scope*: Article 4 applies to any organization using AI in a professional context, regardless of size - the company whose employees “use ChatGPT on their own” is precisely the company in scope, because shadow AI use in a work context is the employer's responsibility.

Second, *proportionality*: the required literacy scales with role and risk; nobody must turn bookkeepers into ML engineers, but everyone touching AI must understand its fundamentals, limits and failure modes - and the organization must be able to demonstrate the measures it has taken, which makes this a documentation obligation as much as a training one.

Third, *timing*: firms that skip the literacy baseline will struggle with the deeper obligations arriving through 2026–2028, from human oversight to documentation, because untrained people cannot perform documented oversight - and regulators are expected to treat a documented literacy gap as an aggravating factor when investigating other AI Act breaches.<sup>60</sup>

Whether in Europe or elsewhere, whether to comply with regulation or “simply” to transform the business for the AI era, choosing a training provider to implement AI literacy falls to HR - or, in many mid-sized firms, to a managing director handling personnel matters alongside the day job, since fewer than half of small firms have a dedicated HR department at all.<sup>61</sup> Those making the choice rarely possess the skills to probe the AI expertise of their training vendors.<sup>61, 62, 63, 64</sup> *HR is thus simultaneously the buyer of AI literacy and its most underserved customer.*

The vendor side compounds the problem. Earlier in this study we discussed the rise of the incompetent AI expert – the consultant or keynote speaker whose entire specialization, on inquiry, turns out to be “ChatGPT” and “using AI apps.” Even a Big Four brand is no guarantee of knowledge,<sup>65</sup> as described in Chapter 2 of this Study.

### *Case Study: The Chief Learning Officer*

The structural answer to the training problem is to give it an owner with the right reporting line. Lauterbach, drawing on three AI implementations at mid-sized industrial champions in 2025-26, argues that the Chief Learning Officer (CLO) role must be rebuilt for the AI era: not a training manager with a new title inside HR, but a capability architect positioned at the intersection of human capability and technical infrastructure - which makes the natural partnership the CTO, not only the CHRO.<sup>66</sup> The logic is structural: the systems that now capture, store and deploy knowledge - RAG (retrieval-augmented generation) databases, avatar platforms, knowledge graphs - are technology assets. Without CTO alignment, the CLO builds learning programs disconnected from the firm's intelligence architecture; without CLO alignment, the CTO builds systems with no human knowledge inside them.

The role's four mandates map directly onto the gaps this section has documented. First, prioritize strategic knowledge using a Knowledge Risk Matrix - whose knowledge, if lost tomorrow, would most damage the firm's competitive position? - answering the knowledge-capture blind spot. Second, build the avatar program jointly with the CTO, with the CLO owning knowledge elicitation, consent and validation - answering the governance question of who owns the expert's avatar, with the crucial guardrail that the expert retains review rights and a defined sunset protocol governs the avatar after departure. Third, translate strategy into a capability roadmap each quarter, so people are not trained for yesterday's strategy - answering the vendor-noise problem, because a firm that knows its capability gaps can finally judge what a training provider must deliver. Fourth, protect knowledge sovereignty: ensure proprietary expertise is captured before it retires, structured so AI can reason with it, and governed so it does not leak through shadow AI use.

The whole program should be framed explicitly as capability amplification, not workforce reduction. For an SME, the appeal of this design is that it is small and fast: a weekly 30-minute CLO-CTO sync, a monthly strategy alignment, a quarterly knowledge audit, and a 12-month sequence from Knowledge Risk Matrix to a first knowledge-sovereignty report to the board.

### **Frontier 4: Culture**

Every AI implementation is, at bottom, a change management project - and the evidence says the change is the hard part. BCG's AI Radar survey of more than 1,400 executives supports what the firm distills as the 10-20-70 rule: only 10% of AI success depends on algorithms and 20% on technology and data, while 70% depends on people, processes and cultural transformation - and BCG's analysis of the CHRO role states plainly that the greatest barriers to realizing AI value are not technological.<sup>67</sup> The problem, in other words, is not technical but a matter of leadership: whether leaders build the trust, redesign the workflows and model the behaviors that let people actually change how they work.

For SMEs the change burden is heavier still, because the change capability is weakest precisely where it is needed: in German SMEs, change management ranks

among the least mature of all seventeen HR fields,<sup>61</sup> and in the OECD's seven-country SME survey, most firms hold neutral or positive attitudes toward generative AI yet only a minority of adopters provide training or internal guidelines – adoption without accompaniment, which is where cultural resistance and shadow use incubate.<sup>14</sup> A quarter of French SME managers describing themselves as “blocked” despite understanding the stakes<sup>62</sup> is not a technology gap; it is a leadership and change gap.

Within this change project, culture is HR's territory - shared, of course, with the executive leadership team and the board - and here HR faces a challenge it has not yet recognized. In mid-sized companies and their larger peers alike, HR owns culture surveys and culture development. Yet HR is today almost entirely oblivious to the fact that, in the case of LLMs, different models exhibit measurably different personalities: consistent patterns in tone, agreeableness, risk appetite and communication style that research has begun to quantify - Google DeepMind researchers have shown that personality traits in LLM outputs can be measured with adapted psychometric instruments, are reliable and valid under certain conditions, and can be shaped to desired profiles<sup>68</sup> - and that shape how humans collaborate with them. edutAlnment confirmed these findings with additional sources.<sup>69</sup>

A company that deploys one model in customer service, another in engineering and a third in finance has, without noticing, hired three colleagues with three different temperaments. When teams become blends of humans and agents, the “culture” HR is charged with developing includes the machine personalities in the room: which model's assertiveness suits a compliance review, whose sycophancy might amplify a manager's bad idea rather than challenge it, how an agent's communication style lands with an apprentice versus a thirty-year veteran. A function that measures human culture in annual surveys has no instruments at all for the culture of a blended human-AI team - no baseline, no metrics, no vocabulary - even as such teams become the operating norm. Closing that gap does not require HR to become technical; it requires HR to lead: to extend its oldest competencies - observing how people work together, naming what helps and what corrodes, and holding leaders accountable for the climate they create - to teams in which some of the colleagues are models.

### ***Recommendations for HR on AI Literacy:***

- First, prioritize HR itself as the first recipient of AI training: the function that must design the hybrid workforce cannot remain its least literate member;
- Second, run HR as a business – meaning HR fully understands the company's strategic priorities and competitive position, and can translate AI workforce decisions into business terms rather than administering them;
- Third, in larger companies, establish the Chief Learning Officer role - a capability architect aligned with the CTO as well as the CHRO - to professionalize the selection of training vendors and partners and to own the knowledge-capture agenda;
- Fourth, use HR as one of the first functions to be automated with AI - deliberately, as a showcase of both potential and risk: a function that has felt automation bias, cost surprises and model-personality quirks in its own processes will regulate them more wisely for everyone else.

## Implementing AI in SMEs: A Practical Roadmap

There is no such thing as an AI strategy in the absence of a business strategy. A company that has not made the hard decisions - which markets it will fight for and which it will exit, which capabilities are core and which are commodity, what it will stop doing, who will own the business in five years - cannot decide what AI is *for*. It can only decide what AI is *like*, which is how eighty percent of pilots come to exist: as imitations of a competitor's press release. As César Hidalgo puts it in his work on knowledge and organizations, deliberation is necessary, but at some point someone must cut the cake - and the ability to make that decision well is judgment, the rarest and most valuable organizational resource under uncertainty.<sup>1</sup>

Only downstream of those hard decisions does an AI strategy make sense: a deliberate answer to where AI creates value in *this* business - which workflows, which decisions, which knowledge - including the honest option of *not yet* in areas where the proof is missing. And only downstream of the AI strategy does AI governance become what it should be: not a compliance checklist filed to satisfy auditors, but the discipline of clarity - the instrument by which an organization navigates according to what it believes its purpose, market, growth potential, and operational constraints to be. In the grey areas that AI regulation inevitably leaves, the governing question is not where a company lands but whether it is there on purpose: able to defend the data it chose not to use, the automation it chose not to deploy, why it stopped where it stopped. And here is the dependency this study exists to expose: you cannot be somewhere on purpose in a landscape you cannot read. Every element of deliberate positioning demands AI literacy. In Chapter 2 we have already discussed the trap of *willful blindness*: if there is information you could have known, and should have known, but contrived not to know, the law treats you as though you knew it. AI literacy removes the excuse of not having looked.

The sequence - strategy, then AI strategy including writing down core beliefs, then governance, each enforced by AI literacy - is not corporate theater for small companies. It is the foundational work of understanding their comparative advantage. Hidalgo's observation about scarcity applies with precision: with constrained resources, a small team of genuinely skilled people who own the problem completely will consistently outperform a larger, better-funded effort with ownership distributed across many functions and meetings. Scarcity without skill is just poverty; but in organizations that have the skill, budget constraints produce exactly the focused, personally-invested dynamic that solves hard problems.<sup>1</sup> An SME cannot out-spend a large international enterprise on AI. It can out-decide one.

In practice, the roadmap looks as follows.

*Strategy first, with the hard decisions* - markets, capabilities, ownership, succession - because a company that does not know where it is going cannot know what AI is for.

*AI strategy and governance second*, grounded in the core beliefs and values of the company, which dictate its technology choices. Deciding which problems AI must solve first is imperative: capturing and automating knowledge as the workforce

retires; building avatars that carry specific skills to train vendors and partners in multiple languages, if growth plans demand working within an ecosystem rather than leveraging internal capabilities alone. Building substantial technical AI expertise internally may open the door to more sophisticated open-weight tools and keep costs in view. If cybersecurity dominates as a principle, working with AI vendors specialized on highly-regulated industries may be unavoidable. Building a powerful internal AI technology teams might guarantee that the company can keep segments of its value chain insourced rather than outsourcing them to cheaper labor markets, if business values control over its production and delivery more than anything else. Strategic clarity ensures workflow archaeology before tool procurement, the buy-versus-build diagnostic before the contract, and the honest option of *not yet* where proof is missing.

*Protect the human skills:* price experience as scarce capital, capture the knowledge before it retires, give juniors more reps rather than fewer, and reserve for humans what the machine cannot do - the judgment, the client relationship, the art to hire best talent.

*Build organizational elasticity, within the Orchestration Tax:* humans and agents as one flexible system, oversight never scaled beyond human capacity, and every consequential loop kept auditable.

*Lead and shape the local ecosystem rather than waiting for it.* An SME cannot buy a research department, but it can borrow one: the evidence is unambiguous that ecosystems accelerate AI. One recent survey-based study found SMEs collaborating with universities and research centers were six times more likely to adopt AI,<sup>19</sup> and analyses of European SME data confirm that a supportive business environment - collaboration with educational institutions, research organizations and business partners, plus access to skills and finance - is itself a driver of adoption, interacting synergistically with a firm's internal capabilities.<sup>70</sup> Four moves are available to almost any mid-sized firm:

First, *partner with the local university or university of applied sciences:* thesis projects, knowledge-transfer partnerships, and shared-facility programs bring current methods into the firm at a fraction of consultancy rates, and research shows higher-education institutions support SME adoption both actively through collaboration and passively through knowledge spillover, especially in the early stages.<sup>71</sup>

Second, *hire interns directly from schools and colleges to spot emerging technical talent early* - a mid-sized firm that offers a sixteen-year-old real problems and real data can out-recruit a corporation offering a rotation program, and it builds the junior pipeline this study's workforce chapter shows is thinning.

Third, *involve startups in solving practical problems:* a startup needs a reference customer and real-world data as much as the SME needs speed, and a pilot scoped around one concrete workflow is cheaper than any enterprise software contract - with the one-primary-vendor discipline of this study applied to what goes into production, so ecosystem experimentation does not become frankensteinization.

Fourth, *treat the collaboration as an IP opportunity, not just a service purchase*: jointly developed models, fine-tuned weights, datasets and workflow designs are assets, and ownership, licensing and publication rights should be settled in the agreement at the start – following established industry practice for collaborative R&D and the same contractual discipline this study recommends for expert avatars.

Handled this way, *the ecosystem is not a substitute for internal literacy but its accelerant*. Policymakers have built the on-ramps - from European Digital Innovation Hubs to national AI voucher schemes and the public-private partnerships the G7 SME AI Adoption Blueprint calls for - but on-ramps carry only the firms that drive onto them.<sup>72</sup>

Fifth, *educate the whole company*: how LLMs and agents actually work versus how humans think; the cognitive science of working with AI, so productivity is not purchased with atrophy; data, privacy and cybersecurity hygiene in the age of perfect phishing; the economics of tokens and the regulatory floor; and how to start, measure, and kill an AI project.

### Sources to Chapter 3\_Section 4

1. César Hidalgo, *The Infinite Alphabet* (2026); the passage is discussed in Anastassia Lauterbach Interview titled *Judgment Is the Rarest and Most Valuable Organizational Resource*. Leading Through Disruption, 4 June 2026, Available from <https://www.linkedin.com/pulse/judgment-rarest-organizational-resource-dr-anastassia-lauterbach-gnxae/?trackingId=u6Pfg%2FHVTL2xSP%2B9xcGsYw%3D%3D> (cited Aug 25 2026)
2. Lauterbach A., *AI Literacy in SMEs*. Original Survey, n=915 (with 276 full responses including open questions); percentages refer to the respondents answering each question. In October and November 2026, the Survey will be complemented with further data from Europe. Update on Dec 1<sup>st</sup> 2026, available from [www.aiedutainment.ai](http://www.aiedutainment.ai)
3. Bick A. et al., *Mind the Gap: AI Adoption in Europe and the US*, Working Paper 34995. March 2026. NBER. National Bureau of Economic Research. Available from <https://www.nber.org/papers/w34995> (cited Aug 25 2026). Surveys of 5,000+ workers in the US and six European countries (2025–2026) plus a firm survey covering 32 European countries. US firm-level AI integration is estimated at 34% versus an EU-wide average of 20%; 43% of US workers report using generative AI at work in early 2026, against 26% (Italy) to 36% (UK) in Europe. The paper links higher AI adoption to faster industry-level productivity growth, against the backdrop of US productivity growing almost 90% since 1995 versus roughly 30% in the euro area.
4. *Accenture Report: European Firms Must Accelerate AI Adoption to Close Productivity Gap*. June 26 2025. Available from

<https://newsroom.accenture.com/news/2025/accenture-report-european-firms-must-accelerate-ai-adoption-to-close-productivity-gap> (cited Aug 25 2026)

5. Eurostat, *20% of EU Enterprises Use AI Technologies*. Dec, 11 2025. Available from <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20251211-2>; Eurostat. Use of Artificial intelligence in Enterprises. Dec 2025. Available from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use of artificial intelligence in enterprises](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises); Eurostat. *Artificial intelligence By Size Class of Enterprise*. Update June 15 2026. Available from [https://ec.europa.eu/eurostat/databrowser/view/isoc\\_eb\\_ai/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/isoc_eb_ai/default/table?lang=en) (all cited Aug 25 2026). ICT Usage in Enterprises (2024–2025): across all sectors, 87.4% of European SMEs do not use AI, with country-level adoption ranging from over 35% in Denmark, Finland and Sweden to under 10% in several southern and eastern member states.
6. Draghi, M. *The Future of European Competitiveness*, European Commission. Sept, 2024. Available from [https://commission.europa.eu/topics/competitiveness/draghi-report\\_en](https://commission.europa.eu/topics/competitiveness/draghi-report_en) (cited Aug 25 2026)
7. Parasuraman, R. & Manzey, D.H. (2010), *Complacency and Bias in Human Use of Automation: An Attentional Integration*, Human Factors 52(3) - the integrative review across aviation, medicine and military contexts, showing the tendency to skip independent verification is especially pronounced under cognitive load, and that complacency appears in both naive and expert users and cannot be overcome with simple practice. Available from [https://www.researchgate.net/publication/47792928 Complacency and Bias in Human Use of Automation An Attentional Integration](https://www.researchgate.net/publication/47792928_Complacency_and_Bias_in_Human_Use_of_Automation_An_Attentional_Integration) (cited Aug 25 2026)
8. HBS. *Is AI Making Your Team Lazy?* Available from <https://aiinstitute.hbs.edu/is-ai-making-your-team-lazy/> (cited Aug 25 2026)
9. Kosmyna, N. et al. (2025), *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*, MIT Media Lab, arXiv:2506.08872. Complement with Lee, H.-P. et al. (2025) Available from [https://www.researchgate.net/publication/392560878 Your Brain on ChatGPT Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task](https://www.researchgate.net/publication/392560878_Your_Brain_on_ChatGPT_Accumulation_of_Cognitive_Debt_when_Using_an_AI_Assistant_for_Essay_Writing_Task) (cited Aug 25 2026)
10. UpGuard, *The State of Shadow AI*. Nov 2025. Available from <https://www.upguard.com/resources/the-state-of-shadow-ai> (cited Aug 25 2026). The survey analyzed responses from 500 security leaders and 1,000 employees across the globe. 81% of employees and 88% of security leaders report using unapproved AI tools. 45% of workers find workarounds to access

blocked applications. Even with 40% of employees recalling AI training, 40% still use unapproved tools on a daily basis. BlackFog. *Shadow AI Threat Grows Inside Enterprises as BlackFog Research Finds 60% of Employees Would Take Risks to Meet Deadlines*. Apr 27 2026. Available from <https://www.blackfog.com/blackfog-research-shadow-ai-threat-grows/> (cited Aug 25 2026)

11. *AI in Business*. Office of Advocacy. Census Bureau's Business Trends and Outlook Survey, Sept. 2023 – Aug 2025. Available from <https://advocacy.sba.gov/wp-content/uploads/2025/09/Research-Spotlight-AI-in-Business-Small-Firms-Closing-In-092425.pdf> (cited Aug 25 2026)
12. Mavvrik. *2026 State of AI Cost Governance Report*. Available from (upon registration): <https://www.mavvrik.ai/state-of-ai-cost-governance-report/#2026report-form> (cited Aug 25 2026)
13. Brynjolfsson, E., Rock, D. & Syverson, C. (2021), *The Productivity J-Curve: How Intangibles Complement General Purpose Technologies*, American Economic Journal: Macroeconomics 13(1), 333–372. Available from <https://www.aeaweb.org/articles?id=10.1257/mac.20180386> (cited Aug 25 2026). The canonical academic statement supporting edutAlnment findings: transformative technologies require complementary intangible investments – process redesign, human capital, new business models – that are largely invisible in standard accounting, so measured costs understate true costs and measured productivity initially understates true progress.
14. OECD. *Generative AI and the SME Workforce. New Survey Evidence*. Nov 5 2025. Available from [https://www.oecd.org/en/publications/generative-ai-and-the-sme-workforce\\_2d08b99d-en.html](https://www.oecd.org/en/publications/generative-ai-and-the-sme-workforce_2d08b99d-en.html) (cited Aug 25 2026)
15. Department for Science, Innovation and Technology (DSIT), *AI Adoption Research*. IFF Research and Technopolis. Jan 28 2026, updated Feb 13 2026. Available from <https://www.gov.uk/government/publications/ai-adoption-research/ai-adoption-research> (cited Oct 11 2026). 16% of UK businesses use AI (micro 14%, mid-sized 23%, large 36%); limited AI skills, expertise or knowledge is the second most-cited barrier (60%). Fieldwork February-May 2025.
16. Office for National Statistics (ONS), *Artificial Intelligence in UK Businesses: 2023 to 2026*. July 20 2026. Available from <https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/articles/artificialintelligenceinukbusinesses/2023to2026> (cited Oct 11 2026). Business Insights and Conditions Survey (BICS), Wave 159: AI use among firms with 10+ employees rose from around 12% to around 35% since late 2023; a lack of expertise is one of the most-cited barriers; only 11% of firms

with 10+ employees report that more than half of their workforce has received AI-related training.

17. *EU Artificial Intelligence Act. Article 4: AI Literacy*. Feb. 2 2025. Available from <https://artificialintelligenceact.eu/article/4/> (cited Aug 25 2026)
18. Cuatrecasas, *Digital Omnibus on AI: How Does It Impact Employment Relations?* July 25 2026. Available from <https://www.cuatrecasas.com/en/global/labor-and-employment/art/digital-omnibus-on-ai-how-does-it-impact-employment-relations/>; Van Eecke P., Regenhardt B., *Digital AI Omnibus Delays Key Deadlines, Introduces New Rules*. Cooley. Aug 3 2026. Available from <https://cdp.cooley.com/digital-ai-omnibus-delays-key-deadlines-introduces-new-rules/>; Interface, *An Operational Framework for AI Literacy in the Workplace*. Policy Brief. Aug 25 2026. Available from <https://www.interface-eu.org/index.php/publications/ai-literacy-workplace-framework> (all cited Oct 11 2026). The Digital Omnibus on AI entered into force on July 27 2026.
19. Sánchez E. et al. (2025), *Artificial Intelligence Adoption in SMEs: Survey Based on TOE–DOI Framework, Primary Methodology and Challenges*, MDPI, Applied Sciences 15(12): 6465. Available from <https://www.mdpi.com/2076-3417/15/12/6465> (cited Aug 15 2026) SMEs collaborating with universities and research centers showed a six-fold increase in the likelihood of AI adoption.
20. OECD, *AI Adoption by Small and Medium-Sized Enterprises: OECD Discussion Paper for the G7*. Dec 9 2025. Available from [https://www.oecd.org/en/publications/2025/12/ai-adoption-by-small-and-medium-sized-enterprises\\_9c48eae6.html](https://www.oecd.org/en/publications/2025/12/ai-adoption-by-small-and-medium-sized-enterprises_9c48eae6.html) (cited Aug 25 2026)
21. OECD, *Empowering SMEs in the Age of AI: The 2026 OECD D4SME Survey* (OECD Centre for Entrepreneurship, SMEs, Regions and Cities, 2026). 2,018 responses across 12 countries; 61% of SMEs report using at least one AI-enabled application; 76% of AI-using firms classified as “AI novices.” Available from [https://www.oecd.org/en/publications/empowering-smes-in-the-age-of-ai\\_bf5a9816-en.html](https://www.oecd.org/en/publications/empowering-smes-in-the-age-of-ai_bf5a9816-en.html) (cited Aug 25 2026)
22. Grundy A., Breaux C., Khatiwoda B., *Large Firms With at Least 20 Employees Biggest AI Users*. U.S. Census Bureau. May 26 2026. Available from <https://census.gov/library/stories/2026/05/ai-use-businesses.html> (cited Oct 11 2026). Business Trends and Outlook Survey (BTOS): national AI use rate of 19.8% as of May 3 2026, within a 17-20% range since December 2025; 37% among firms with 250+ employees and 32% among firms with 100-249 employees; use rose among firms with 20+ employees but did not change significantly among firms with fewer than 20.

23. U.S. Chamber of Commerce Technology Engagement Center (C\_TEC) and Teneo, *Empowering Small Business: The Impact of Technology on U.S. Small Business 2026*. Oct 2 2026. Available from <https://www.uschamber.com/technology/2026-empowering-small-business-report> (cited Oct 11 2026). Survey of 3,732 firms with fewer than 250 employees (fieldwork July 23-Aug 6 2026): 66% of small businesses use AI, up 8 points from 2025; 95% of small businesses using AI are upskilling employees; six in ten worry about a patchwork of state privacy and AI laws.
24. Federal Reserve Banks, *2026 Report on Employer Firms: Findings from the 2025 Small Business Credit Survey*. Mar 3 2026. Available from <https://www.fedsmallbusiness.org/2026-report-on-employer-firms> (cited Oct 11 2026). 46% of employer firms currently use AI and 15% plan to within 12 months; 7% of users have fully integrated AI; among firms planning to adopt, 37% cite the time needed to implement AI or train employees as a challenge. Convenience sample.
25. Office for National Statistics (ONS), *Business Insights and Impact on the UK Economy: 8 October 2026*. Oct 8 2026. Available from <https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/bulletins/businessinsightsandimpactontheukeconomy/8october2026> (cited Oct 11 2026). BICS Wave 165: 30% of UK businesses used at least one AI technology in September 2026; 51% of businesses with 100 or more employees against 29% of businesses with fewer than 10 employees.
26. Original Source: Infocomm Media Development Authority (IMDA), *Singapore Digital Economy Report 2026*. Oct 5 2026. Used source: TNGlobal, *Singapore's Digital Economy Reaches 19.3% of GDP as AI Adoption Deepens: IMDA*. Oct 5 2026. Available from <https://technode.global/2026/10/05/singapores-digital-economy-reaches-19-3-of-gdp-as-ai-adoption-deepens-imda/> (cited Oct 11 2026). SME AI adoption rose to 23.4% from 14.5% a year earlier; large local firms to 70.4% from 62.5%.
27. European Investment Bank, *European Firms Show Resilience, Invest in Green Transition and Match US Companies in Adopting AI Technologies, New EIB Survey Shows*. EIB Investment Survey 2025. Oct 14 2025. Available from <https://www.eib.org/en/press/all/2025-383-european-firms-show-resilience-invest-in-green-transition-and-match-us-companies-in-adopting-ai-technologies-new-eib-survey-shows> (cited Oct 11 2026). 37% of EU firms use generative AI against 36% in the US; only 55% of EU firms using AI apply it in more than two business activities, against 81% in the US.
28. Lewis Silkin, *AI Judgment Day on the Horizon While UK Lawmakers Play Catch Up*. Sept 14 2026. Available from <https://www.lewissilkin.com/insights/2026/09/14/ai-judgment-day-on-the->

- [horizon-while-uk-lawmakers-play-catch-up](#); Horwood P., *Government Announces Free AI Training for 10 Million Workers by 2030*. Computing. Jan 28 2026. Available from <https://www.computing.co.uk/news/2026/government/govt-plan-ai-training-10-million-2030> (both cited Oct 11 2026)
29. Deloitte UK, *British Workers Spend Nearly £1bn of Their Own Money on GenAI for Work*. Press release on the Deloitte GenAI Workforce Survey, conducted with Ipsos. Sept 16 2026. Available from <https://www.deloitte.com/uk/en/about/press-room/british-workers-spend-one-billion-pounds-of-their-own-money-on-gen-ai-for-work.html> (cited Oct 11 2026). Survey of 25,000 UK working adults aged 18-70: 63% knowingly use GenAI for work; 31% of users use it without their employer's knowledge; 17% pay for at least one tool themselves (an estimated £958m a year); around half of users have received no formal training.
  30. Levi, Y., *Tank vs. Drone: Rethinking Cybersecurity Strategy in the Age of AI*, Whitepaper. Levi synthesizes the G7 Cyber Expert Group Statement on AI and Cybersecurity (Oct 2025), ENISA Threat Landscape 2025, WEF Global Cybersecurity Outlook 2026, NIST AI 100-2e2025 (adversarial ML taxonomy), MITRE ATT&CK and ATLAS, and CISA/NSA/NCSC secure-AI guidance. Available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Aug 25 2026)
  31. *2026 Data Breach Investigations Report*. Verizon. Available from <https://www.verizon.com/business/resources/T1ae/reports/2026-dbir-data-breach-investigations-report.pdf> (cited Aug 25 2026)
  32. Bourtole, L. et al., *Machine Unlearning*. Dec. 14 2020. 42<sup>nd</sup> IEEE Symposium of Security and Privacy. Available from <https://arxiv.org/pdf/1912.03817> (cited Aug 25 2026)
  33. IBM, *Cost of a Data Breach Report 2026*. Available from <https://www.ibm.com/reports/data-breach> (cited Aug 25 2026)
  34. OWASP Foundation (2025), *OWASP Top 10 for Large Language Model Applications*. <https://owasp.org/www-project-top-10-for-large-language-model-applications/>
  35. Mozilla Foundation, *The state of Open Source AI*. July 2026. Available from <https://stateofopensource.ai/> ; OpenAI, *Introducing gpt-oss*. Aug 5 2025. Available from <https://openai.com/index/introducing-gpt-oss/> (cited Aug 25 2026)
  36. HuggingFace, *Statue of Open Source on Hugging Face: Spring 2026*. Mar 17 2026. Available from <https://huggingface.co/blog/huggingface/state-of-os-hf-spring-2026>; *The State of Open-Source LLMs in 2026*. Aug 8 2026. Available from <https://blog.redlinesoft.net/posts/state-of-open-source-llms-2026/> (both cited Aug 25 2026)

37. Lauterbach A., *Trojanische Verhältnisse? In China Im Blickpunkt des 21. Jahrhunderts*. Loitsch t. (Hrsg.), SpringerGabler, 2019 (2<sup>nd</sup> edition 2021) (cited Aug 25 2026)
38. Apotheker J. et al., *As AI Investments Surge, CEOs Take the Lead*. BCG. Jan 15 2026. Available from <https://www.bcg.com/publications/2026/as-ai-investments-surge-ceos-take-the-lead> (cited Aug 25 2026)
39. Rowinski D., *How AI Is Driving Revenue, Cutting Costs and Boosting Productivity for Every Industry in 2026*. Nvidia. Mar 9 2026. Available from <https://blogs.nvidia.com/blog/state-of-ai-report-2026/> ; *Ramp AI index*. July 2026. Available from <https://ramp.com/data> (both cited Aug 25 2026)
40. *Global AI Pulse Q2 2026*. KPMG. Available from <https://kpmg.com/xx/en/our-insights/ai-and-technology/ai-pulse.html>; PWC, *Leaving Through Uncertainty in the Age of AI. PWC's 29th Global CEO Survey*. Available from <https://www.pwc.com/gx/en/ceo-survey/2026/pwc-ceo-survey-2026.pdf> (both cited Aug 25 2026)
41. Hadby J. B., *Reducing AI Token Costs*. How to Cut LLM Spend Without Cutting Value. July 2026. Internal AI. Available from <https://internal.ai/reducing-ai-token-costs> (cited Aug 25 2026)
42. *AI Costs and the Future of Tech Finance* with Carmen Li, Episode 78, podcast AI Snacks with Romy&Roby.
43. Microsoft, *Work Trend Index Annual Report: The Year the Frontier Firm Is Born*, April 2025. Available from <https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born> (cited Aug 25 2026)
44. Korn Ferry (2025), *CHRO Survey 2025*. Survey of 756 HR leaders across 50+ countries: only 5% of HR teams describe themselves as well-prepared to integrate AI, none as fully prepared; 40% of CHROs cite insufficient AI-related knowledge and skills within HR teams as the biggest obstacle to integrating AI into talent management. <https://www.kornferry.com/insights/featured-topics/leadership/role-of-the-chro> - For the ownership claim: Bessemer Venture Partners (2025), *The AI upskilling guide for executives: responsibility for upskilling* "is typically handed off to the Chief Human Resources Officer." Available from <https://www.bvp.com/atlas/the-ai-upskilling-guide-for-executives> (cited Aug 25 2026)
45. Gartner, *Gartner Research Finds Only 8% of HR Leaders Believe Their Managers Have the Skills to Effectively Use AI*. Oct 8 2025. Available from <https://www.gartner.com/en/newsroom/press-releases/2025-10-08-gartner-research-finds-only-8-percent-of-hr-leaders-believe-their-managers-have-the-skills-to-effectively-use-ai> (cited Aug 25 2026)

46. Osmani A., *The Orchestration Tax*, May 24 2026. Available from <https://addyosmani.com/blog/orchestration-tax/> (cited Aug 25 2026)
47. Samanta N., *The Growing Challenge of Auditing Agentic AI*. ISACA. Sept 2 2025. Available from <https://www.isaca.org/resources/news-and-trends/industry-news/2025/the-growing-challenge-of-auditing-agentic-ai>; IBM, *Building trustworthy Ai Agents for Compliance: The Challenge of Auditability and Explainability*. Available from <https://www.ibm.com/think/insights/building-trustworthy-ai-agents-compliance-auditability-explainability>, Osmond M. et al. *Mind the Gap: How the Technical Mechanisms of Agentic AI Outpace Global Legal Frameworks*. Mar 31 2026. Available from <https://arxiv.org/pdf/2603.27075> (all cited Aug 25 2026)
48. Bratt A., *Who Owns your Digital Twin?* Emerging Employment Law Issues for Employers. June 30 2026. Gowling WLG. Available from <https://gowlingwlg.com/en/insights-resources/articles/2026/emerging-employment-law-issues-for-employers> (cited Aug 25 2026)
49. *AI Workforce with Sam Huber*. Episode 88, podcast AI Snacks with Romy&Roby (cited Aug 25 2026)
50. Brill J, Wunker S., *AI and the Octopus Organization: Building the Superintelligent Firm*. Second Edition. Menlo Park Books, Sept 2025. ISBN B0FQQ4SSLL8; Brill J., CIO. *The Octopus Playbook for AI: What Nature's Smartest Cephalopod Can Teach Leaders About AI*. Dec 17 2025. Available from <https://jonathanbrill.com/media/cio-the-octopus-playbook-ai/> (both cited Aug 25 2026)
51. Deloitte, *The \$9 Trillion Knowledge Exodus: How Organizations Can Turn Baby Boomer Retirements Into A Competitive Advantage*. June 18 2026. Available from <https://www.deloitte.com/us/en/insights/topics/talent/knowledge-management-plan.html> (cited Aug 25 2026)
52. Brynjolfsson, E., et al., *Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence*, Stanford Digital Economy Lab. Aug 12 2026. Available from <https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/> (cited Aug 25 2026)
53. European Commission, *New Commission Guidance Will Make It Easier to Transfer Ownership of SMEs*. June 23 2026. Available from [https://single-market-economy.ec.europa.eu/news/new-commission-guidance-will-make-it-easier-transfer-ownership-smes-2026-06-23\\_en](https://single-market-economy.ec.europa.eu/news/new-commission-guidance-will-make-it-easier-transfer-ownership-smes-2026-06-23_en) (cited Oct 11 2026)
54. Choi J., *No Heirs, No Buyers: Korea's SME Succession Dilemma*. The Korea Herald, republished by Asia News Network. June 5 2026. Available from

- <https://asianews.network/no-heirs-no-buyers-koreas-sme-succession-dilemma/> (cited Oct 11 2026). Based on Ministry of SMEs and Startups data as of end-2025: about one-third of Korean SMEs have owners aged 60 or older; 28.6% of these owners have no successor.
55. Original Source: McKinsey Institute for Economic Mobility, *The Great Ownership Transfer: A New Era of Business Stewardship*. 2026. Used source: Delozier J., *Boomer Retirements Could Trigger Record Small-Business Transfers*. HousingWire. Mar 2 2026. Available from <https://www.housingwire.com/articles/baby-boomer-retirement-small-business-transfers/> (cited Oct 11 2026). By 2035, about 6 million small and medium-sized businesses will face ownership transitions as their owners retire; effective transitions could keep up to 12 million jobs in place.
  56. Interviews conducted for the AI Literacy Study with Nic Weatherhead, VP of People Operations and Talent Acquisition, The Stepstone Group, and with the GM of Global Talent Acquisition for EMEA & Asia at a large technology group (name withheld), 2026. Transcripts available upon request from [anastassia@aiedutainment.ai](mailto:anastassia@aiedutainment.ai)
  57. Loaiza I., Rigobon R., *The EPOCH of AI: Human-Machine Complementarities at Work*. MIT Sloan Research Paper 7236-24. Oct 1 2025. Available from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5028371](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5028371) (cited Aug 25 2026)
  58. EUR-Lex. *Regulation (EU) 2024/1690 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC)*, available from <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>; *EU Artificial Intelligence Act, Article 3: Definitions*. Feb 2 2025, Available from <https://artificialintelligenceact.eu/article/3/> (both cited Aug 25 2026)
  59. Machin E., *Five Takeaways From the EU Commission's AI Literacy Q&As*. Ropes & Gray. May 21 2025. Available from <https://www.ropesgray.com/en/insights/viewpoints/102kbn5/five-takeaways-from-the-eu-commissions-ai-literacy-qas> (cited Aug 25 2026)
  60. *AI Literacy - Questions & Answers*. Available from <https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers>; *AI Talent, Skills and Literacy*. Available from <https://digital-strategy.ec.europa.eu/en/policies/ai-talent-skills-and-literacy> ; Cooper D. et al., *European Commission Publishes Q&A on AI Literacy*. Inside Privacy. Updates on Developments in Data Privacy and Cybersecurity. Covington. Available from <https://www.insideprivacy.com/artificial-intelligence/european-commission-publishes-qa-on-ai-literacy/> (all cited Aug 25 2026)

61. Knaut, C., Sohr, M., Fischer, M., Drees, J. et al. (2025), *HR-Radar KMU 2025 – Wo steht die Personalarbeit 2025?*, Technische Hochschule Köln, in cooperation with Workfive and Personio. Survey of n = 614 managing directors, HR professionals and staff with HR duties in German SMEs. Only 47% of firms with 1–49 employees have a dedicated HR department (88% at 50–249; 98% at 250–499). Leadership (2.6/5), employee development (2.8/5) and change management (2.9/5) score lowest across 17 HR fields; 50–80% of respondents use no AI in any HR field. [https://www.th-koeln.de/mam/downloads/hrradar2025\\_ergebnisbericht.pdf](https://www.th-koeln.de/mam/downloads/hrradar2025_ergebnisbericht.pdf) (cited Aug 25 2026)
62. Bpifrance Le Lab (2025), *L'IA dans les PME et ETI françaises : une révolution tranquille*, June 2025. Survey of 1,209 SME and mid-cap managers (firms under 10 employees excluded), fielded 15 October–16 December 2024. The "Bloqués" profile – managers who are aware of the stakes but paralysed by a lack of skills, training or support – represents 26% of respondents; 58% consider AI a question of survival within 3–5 years, yet 57% have no AI strategy. [https://lelab.bpifrance.fr/content/download/4745/pdf/2025-06\\_L'IA%20dans%20les%20PME%20et%20ETI%20fran%C3%A7aises\\_Etude%20Bpifrance%20Le%20Lab.pdf](https://lelab.bpifrance.fr/content/download/4745/pdf/2025-06_L'IA%20dans%20les%20PME%20et%20ETI%20fran%C3%A7aises_Etude%20Bpifrance%20Le%20Lab.pdf)
63. Osservatorio Innovazione Digitale nelle PMI, Politecnico di Milano (2026), *Ricerca 2025–2026*, presented 21 May 2026: 76% of Italian SMEs have neither invested nor plan to invest in AI; only 7% have launched structured AI training programmes; only 46% conduct any skills assessment and only 40% periodically draw up training plans beyond mandatory training. <https://www.osservatori.net/comunicato/innovazione-digitale-nelle-pmi/pmi-italiane-innovazione/> — Corroborated by Confindustria (2025), *Indagine Confindustria sul lavoro 2025*: among member firms adopting AI, only 43.7% have taken any HR-side action, and the lack of internal competencies (36.7%) is the most-cited obstacle. <https://www.confindustria.it/pubblicazioni/indagine-confindustria-sul-lavoro-del-2025/> (cited Aug 25 2026)
64. Bessemer Venture Partners, *The AI Upskilling Guide for Executives*. June 24 2025. Available from <https://www.bvp.com/atlas/the-ai-upskilling-guide-for-executives> (cited Aug 25 2026)
65. Original Source: Deloitte Touche Tohmatsu (Australia). 2025. *Targeted Compliance Framework Assurance Review*. Report for the Australian Government Department of Employment and Workplace Relations. Revised version dated 26 September 2025. Used source: Kissin, Ellesheva. 2025. *Deloitte Issues Refund for Error-Ridden Australian Government Report That Used AI*. In *Financial Times*, October 6 2025. Available from: <https://www.ft.com/content/934cc94b-32c4-497e-9718-d87d6a7835ca?syn-25a6b1a6=1>; GPTZero. 2026. *Investigation: Hallucinations in Ernst & Young*

- Report on Loyalty Fraud*. 14 May 2026. Available from <https://gptzero.me/investigations/ey>; Kissin, Ellesheva. 2026. *EY Retracts Study After Researchers Discover AI Hallucinations*. In *Financial Times*, May 15 2026. Available from <https://www.ft.com/content/a61cbcae-95e4-4449-86e1-ef40fb306f4e?syn-25a6b1a6=1>; Esau, Ogale, Cui, *Chasing the Hallucinations: KPMG's AI-Powered Attempt at Redefining Excellence*. Available from <https://gptzero.me/news/investigations-kpmg/>, Bratton, Foley, *KPMG report contained AI hallucinations on benefits of... AI*. In *Financial Times*, June 12 2026, Available from <https://www.ft.com/content/b3828e92-4961-4b39-84f0-c42f33be3c3f?syn-25a6b1a6=1> (all cited July 29 2026)
66. Lauterbach, A. (2026), *Establishing a Learning Officer Aligned with the CTO in a Midsized Business and Augmenting Your Workforce with AI Avatars*, AI Edutainment, June 11 2026. Available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Aug 25 2026)
67. Apotheker J. et al., *From Potential to Profit: Closing the AI Impact Gap*. BCG. Available from [w.bcg.com/publications/2025/closing-the-ai-impact-gap](http://w.bcg.com/publications/2025/closing-the-ai-impact-gap). BCG. Jan 15 2025. Available from <https://www.bcg.com/publications/2025/closing-the-ai-impact-gap>; On the barriers being non-technological and HR's role in "activating the 70%": Beauchene V. et al., *The Reinvention of the CHRO in an AI-Driven Enterprise*. BCG. Feb 2 2026. Available from <https://www.bcg.com/publications/2026/reinvention-of-the-chro-in-an-ai-driven-enterprise>; See also Bedard J., *AI Transformation Is a Workforce Transformation*. BCG. Feb 4 2026. Available from <https://www.bcg.com/publications/2026/ai-transformation-is-a-workforce-transformation> (all cited Aug 25 2026)
68. Serapio-García, G. et al., *Personality Traits in Large Language Models*. Available from <https://arxiv.org/abs/2307.00184> (cited Aug 25 2026)
69. Forthcoming Episode 100 (scheduled for Nov 24 2026), AI Snacks with Romy&Romy, *Interview with Tamilla Triantoro*, transcript available on request (cited Aug 25 2026)
70. Arroyabe M. et al., *Analyzing AI adoption in European SMEs: A Study of Digital Capabilities, Innovation, and External Environment*, ScienceDirect, Dec 2024. Available from <https://www.sciencedirect.com/science/article/pii/S0160791X24002811> (cited Aug 15 2026)
71. Schwaeye J. et al., *The New Normal: The Status Quo of AI Adoption in SMEs*. Aug 13 2024. Tayllor & Francis Online. *Journal of Small Business Management*. Available from <https://www.tandfonline.com/doi/full/10.1080/00472778.2024.2379999>; A. (2024–25), Gisa D., *AI Adoption in SMEs – Empowered by Higher Education Institutions*. PD Thesis. University of Twente. Mar 17 2026. Available from

<https://research.utwente.nl/en/publications/ai-adoption-in-smes-empowered-by-higher-education-institutions/> (both cited Aug 25 2026)

72. *European Digital Innovation Hubs (EDIH) network*. Available from <https://european-digital-innovation-hubs.ec.europa.eu/home>; *The SME AI Adoption Blueprint, G7 2025, Kananaskis*, Available from <https://ised-isde.canada.ca/site/ised/en/sme-ai-adoption-blueprint> (all cited Aug 25 2026)

## Chapter 4. Delivering AI Literacy – A Novel Approach

The first three chapters of this AI Literacy Study established what AI literacy is and how it is different from AI fluency, what it must contain due to the most pressing pain points in studying, individually working with AI, and developing and implementing AI strategies across the nations and industries. It reviewed what the four groups - young people and students, teachers, experienced professionals aged 45 and older, and SMEs know, understand and want to know about AI.

*This last chapter turns to delivery.* It addresses what governments, international organizations, and companies currently propose for introducing AI literacy while identifying one repeating flaw across all of them. They treat AI literacy as a competence that will arrive from somewhere - from a framework, a mandate, a vendor, a national curriculum still to be written - and they say remarkably little about what the people concerned should actually learn. The AI literacy content, however, is the missing cornerstone. Furthermore, there is no clarity on how to apply AI literacy in practice – while working with AI tools and models, building a knowledge management system (or ‘second brain’) for studying or capturing expertise in an organization, or architecting entire human vs. AI systems in the context of elastic organizations.

The chapter then sets out a novel AI literacy program resting on *four principles based on four core beliefs*:

1. *Human agency* is at the center, and shaping of understanding comes before usage of AI tools;
2. *Interdisciplinarity* is a core principle shaping the content, including insights from neuroscience, mathematics, linguistics, computer science and engineering, economics, and literature/arts;
3. *Simplified language* to communicate content to those not skilled in mathematics, computer science, neuroscience or economics, and *edutainment as a principle to educate younger students* about AI technologies;
4. The *method of delivery must be customized* by population: benchmarking for SMEs drowning in an opaque tool market; personalized learning rather than blanket LLM rollouts for schools; and, everywhere, a redoubled investment in the development of minds.

### Implementation of AI Literacy Today

Chapter 3 reviewed a number of AI literacy programs and frameworks; this section broadens this analysis. The supply of AI literacy proposals has grown quickly since 2024. They fall into four groups.

**International frameworks.** UNESCO published two AI competency frameworks in September 2024, one for students and one for teachers. The student framework is organized around four dimensions - a human-centered mindset, the ethics of AI, AI techniques and applications, and AI system design - each at three levels of mastery.<sup>1</sup> The teacher framework adds AI pedagogy and professional development.<sup>2</sup> The EU-OECD AILit Framework, finalized in June 2026 and discussed in Chapter One, sets out nineteen competences across four domains - *Engage, Create, Manage, and Shape*.<sup>3</sup> It will anchor the PISA 2029 assessment of media and AI literacy.

All these documents are advisory. They carry no mandate, no suggestions for funding, and no detailed curriculum. Academic frameworks, from Long and Magerko's seventeen competencies in 2020<sup>4</sup> to Ng<sup>5</sup> and colleagues' four-level model in 2021, follow the same logic of describing the destination rather than the road.

**Legal mandates.** Article 4 of the EU AI Act is the first law in the world to require AI literacy of every organization deploying AI in a professional context.<sup>6</sup> As adopted in 2024 and applicable from 2 February 2025, it required providers and deployers to “ensure, to their best extent, a sufficient level of AI literacy” of their staff. What “sufficient” meant was never benchmarked. The Commission's own Q&A of May 2025 clarified that no format, certificate, or number of hours was prescribed, and that reading a user manual was “likely insufficient.”<sup>7</sup> In July 2026 the obligation was softened further. Regulation (EU) 2026/1744, the Digital Omnibus on AI, replaced Article 4 in full: providers and deployers must now “take measures to support the development of AI literacy,” and the law states explicitly that this “does not require providers or deployers to guarantee any specific level of AI literacy of any individual.”<sup>8</sup> National market surveillance authorities began supervising the provision on 2 August 2026, without a standalone penalty. In eighteen months, the only binding AI literacy obligation in the world went from a *duty of result* to a *duty of effort*. The Commission's companion instrument is a repository of more than forty AI literacy practices submitted by companies and public bodies, published with the disclaimer that replicating them grants no presumption of compliance and implies no endorsement.<sup>9</sup> The repository is useful, but it mostly catalogues formats - e-learning platforms, bootcamps, awareness days - not content.

**National programs.** Chapter 3 reviewed the governments with a vision: Beijing's compulsory AI classes from age six, the UAE's K-12 mandate with ethics built in, Estonia's AI Leap, Singapore's guardrails designed as pedagogy, Chile's good content without an operating plan, and South Korea's AI textbooks stripped of official status after one semester. Two further cases complete the picture. In April 2025 the United States issued Executive Order 14277, “Advancing Artificial Intelligence Education for American Youth,: establishing a White House task force, a Presidential AI Challenge, and a pledge signed by more than sixty organizations to provide resources to K-12 schools over four years.<sup>10</sup> In the US, education is constitutionally a state and local matter, therefore the Order can coordinate and fund but not mandate,

and the content of what students learn is left to the pledging companies. Finland's "Elements of AI," launched in 2018 by the University of Helsinki and Reaktor with the goal of teaching one percent of the Finnish population the basics of AI, remains the best-known example of a content-first program at national scale.<sup>11</sup> It has since reached more than two million registered learners in over 170 countries, was translated into every official EU language during Finland's 2019 Council presidency with the aim of reaching one percent of Europeans, and attracts an unusually diverse audience: roughly 40 percent women and a quarter of learners over the age of 45.<sup>12</sup> Elements of AI teaches machine learning, neural networks, and the philosophy of AI to people with a high-school education and no coding background. It is, in the words of its academic lead, a civics course in AI. The United Kingdom offers another approach. Its AI Skills Boost program, expanded in January 2026 to a target of 10 million workers by 2030, delivers free short courses developed by Accenture, Google, IBM, and Microsoft, checked against a Skills England benchmark and rewarded with a government-backed digital badge.<sup>13</sup> The benchmark lists six foundation skills. They include writing clear instructions for AI tools, using AI tools for routine tasks, and understanding the risks and consequences of using AI.<sup>14</sup>

A government has, in effect, defined the national floor of AI competence as fluency in vendor tools, and asked the vendors to teach these.

**Corporate and EdTech programs.** Eurostat reports that 19.95 percent of EU enterprises with ten or more employees used AI technologies in 2025, up from 13.5 percent a year earlier, and 55 percent of large enterprises; adoption ranges from 42 percent in Denmark to 5.2 percent in Romania.<sup>15</sup> The Article 4 duty to educate the workforce lands on organizations at very different levels of institutional maturity. Improvisation is the norm in how companies approach AI literacy.

The largest supply of AI literacy courses by volume comes from AI makers, EdTech, and the compliance industry that formed around Article 4. Microsoft, Google, OpenAI, Anthropic, and Salesforce each run programs that describe themselves as AI literacy or AI fluency and are, as Chapter 1 showed, largely certifications of usage within their own ecosystems. Since February 2025 a market of Article 4 certificates has emerged in Europe, offering ninety-minute courses and QR-verifiable records from under ten euros per employee.<sup>16</sup> A piece of paper with a compliance stamp is now available at scale, even though the EU AI Act does not demand a certificate. The question is what, if anything, the person holding a piece of paper truly understands about AI and whether this very piece of paper will increase success in AI implementations across industries and geographies.

**A role-based tier framework.** The most serious recent attempt to close the gap between the law and the workplace is a policy brief published on 25 August 2026 by interface, the Berlin think-tank formerly known as Stiftung Neue Verantwortung, authored by Lucia Velasco, Siddhi Pal, and Catherine Schneider.<sup>17</sup> The authors observe that corporate training aims at productivity and teaches people to use tools

well, while the law cares whether operators understand how systems fail and cause harm; that a workforce fluent in prompting can still be illiterate in the sense a regulator means; that the Article 4 training market has no quality floor and every incentive to sell courses that look compliant; and that no published evidence shows that any AI literacy program, corporate or public, does what the law calls for. They cite Lintner's 2024 systematic review, which found sixteen validated AI literacy scales, thirteen of them self-report, most built for students, and none tested across cultures.<sup>18</sup> Their conclusion is that certification should wait until a validated way to measure understanding exists. This AI Literacy Study supports interface's idea.

Their brief's proposal is a three-tier framework keyed to how much and how consequentially a worker interacts with AI, rather than to job title or seniority:

- **L0**, the baseline for everyone under an organization's remit, covers what AI is and is not, which systems the organization uses, and generic risks such as fabricated output, bias, and data exposure, with a record of awareness training as evidence.
- **L1**, for workers required to use AI in their role, adds the capabilities, limits, and failure modes of the specific systems, output verification, prompting, disclosure duties, and escalation.
- **L2**, for persons operating or overseeing high-risk systems, requires documented competence in monitoring, drift and bias detection, and intervention, reviewed annually.

For organizations, the framework becomes a five-step process:

- inventory the systems and the people who touch them;
- assign each person to a tier;
- train to that tier;
- record what was done and why;
- review when systems or roles change.

For policymakers, it recommends adapting instruments that already exist - a free baseline such as Elements of AI, a tiered grid, and a commissioned measurement method - rather than inventing new ones.

Three features of this proposal show both what a good framework can do and where even a good framework stops:

- First, it is *cumulative and role-based, which makes literacy portable between employers and gives SMEs a floor to build on*; that is a genuine improvement on a simple statement of a duty;

- Second, its *content is deferred*. The brief states that the knowledge needed at each tier “can be supplied by existing EU instruments”: DigComp 3.0, the EU digital competence framework whose fifth edition integrates AI across its twenty-one competences, and the AILit framework, which was written for primary and secondary schools.<sup>3, 19</sup> The L0 definition of what AI is refers to the AI Act's own legal definition in Article 3.<sup>16</sup> A worker who can recite a statutory definition and locate a competence statement in DigComp has met the tier and learned nothing about why a model hallucinates, what it costs in a certain context or job environment, or who signs the vendor's forecasts.
- Third, and most consequentially, the tiers stop at the operator. *The person who decides whether to deploy AI at all, which system to buy, for what purpose, and at what cost - the board member, the school director, the owner of a fifty-person firm, the ministry official drafting a national strategy - appears nowhere*, unless he or she happens to operate a high-risk system. Yet this is the most consequential interaction with AI that exists, and Chapter 2 showed that it is precisely the interaction most often performed without understanding. The framework measures depth of interaction with a tool. It does not measure depth of responsibility for the decision to use it.

### **AI Literacy Program - a Novel Approach**

The AI Literacy Study therefore proposes that any framework adopted in Europe or elsewhere add a fourth tier above oversight: a decision tier, applying to everyone who allocates budget to AI, approves its deployment, or sets policy around it.

Its content is the five blocks set out later in this chapter:

- what AI and generative AI are, why machines learn differently from humans;
- realistic risks of AI and reasons why they exist, realistic risk mitigation strategies and related funding;
- AI cost drivers (including costs of a geography and related compliance regulation);
- governance from strategic vision, including precision of vocabulary used to discuss and describe AI;
- application to the concrete company in a concrete industry, an educational institution, or individual use of AI (based on (always updated) AI literacy data serving as a baseline) - including practical mastery of tools and ability to choose among different AI applications and providers.

Its evidence is not a training record but a *demonstrated ability to explain a decision*. Without it, a framework does what every other proposal in this section does: it *locates AI literacy in the people who use AI without focusing on the people who choose it*.

It is remarkable that even large standardization organizations don't sufficiently address AI literacy. The US NIST AI Risk Management Framework never uses the term AI literacy at all, asking only for risk training relevant to duties; ISO/IEC 42001 supplies a practice of recording personnel competence but not its content.<sup>20</sup>

### ***Applied AI Literacy: Tacoma Community College Case Study***

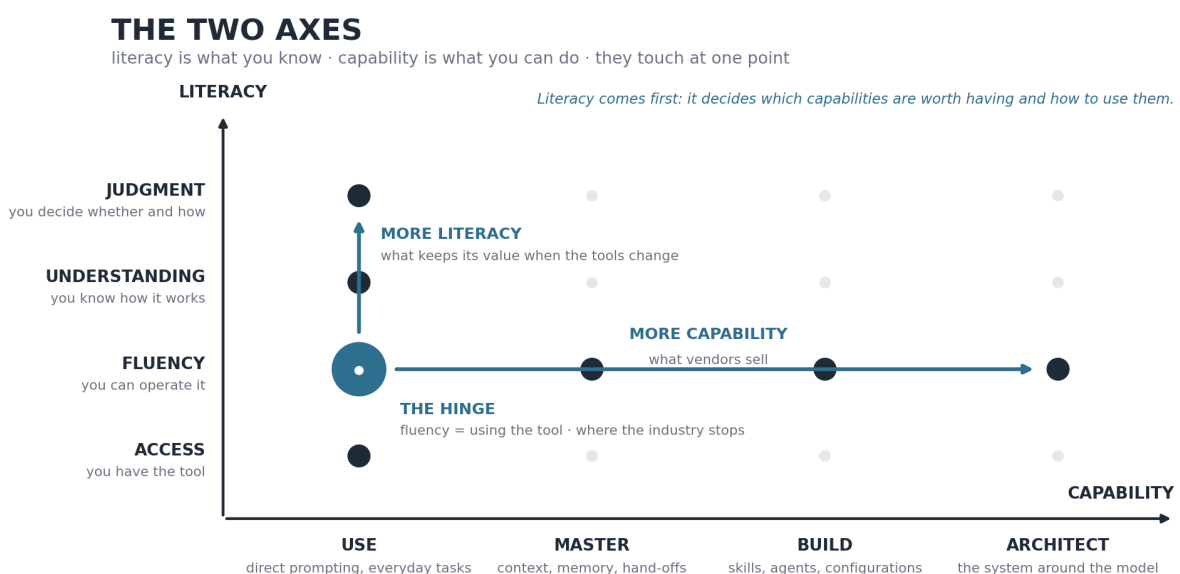
Giorgio Natili's "Beyond Fluency" paper, developed specifically for this Study,<sup>21</sup> details how AI Literacy becomes a real capability to enhance personal and professional development of people. Its recommendations are grounded in socio-economic need, not in academic interest. The starting point was a panel at Tacoma Community College, whose students include first-generation students, veterans, people retraining after layoffs, and professionals updating skills that have begun to age. They were not skeptical about AI. Most were already using a chatbot, mainly as a slightly better search engine, and from very different starting points they converged on one question: "*What should I learn to be competitive, and where do I start?*"

The first answer in the room was the obvious one: start with the tools you already have and learn to use them well. The objection came immediately: the free tool gave wrong answers often enough that people had stopped trusting it, so what was there to learn? The question conflated two things: the tool's capability and the user's. The tool does matter, and paying buys a more capable model with more room to work in, but a subscription buys a more capable instrument, not a more capable operator. The more interesting half of the answer is that even a limited model performs differently depending on the context and instructions it receives. Asking the model to improve your question, a practice called *meta-prompting*, shows this on day one. The technique is fluency; the realization it produces is the first step toward literacy.

That realization is the trap. When AI works badly, an inexperienced user concludes the model is bad. When it works well, the same user concludes it understands more than it does. In both cases, fluency alone offers no way to explain the result. The same error appears at organizational scale: a small team adopting its first automated workflow and a large organization buying a frontier model both tend to choose capability before defining what the system must accomplish, how success will be measured, and what failure costs. The interface hides most of what determines behavior, so product-specific fluency is useful but fragile: interfaces, models, and features change, and what transfers is the ability to understand limits, evaluate output, manage the information a system receives, and decide how much trust to place in it. Fluency is what gets taught because it is what can be measured; understanding is not, and no one who sells a tool is rewarded for teaching people when not to use it.

## Applied AI Literacy Framework

AI literacy comes first and is defined precisely: the ability to recognize and understand what AI is and where its limits are, to evaluate its outputs critically for accuracy and bias, and to use it safely and ethically while keeping a human in a position of judgment and oversight. Natili phases AI Literacy across four stages: *Access, Fluency, Understanding, Judgment*. AI capability is what you can do with the tools, in four stages of its own: *Using, Mastering, Building, Architecting*. *The two axes touch at one point, Fluency, where knowing how to operate a tool and being able to operate it are the same skill. That point is where the industry stops. The order is the argument. Literacy first. Capability read through it.*



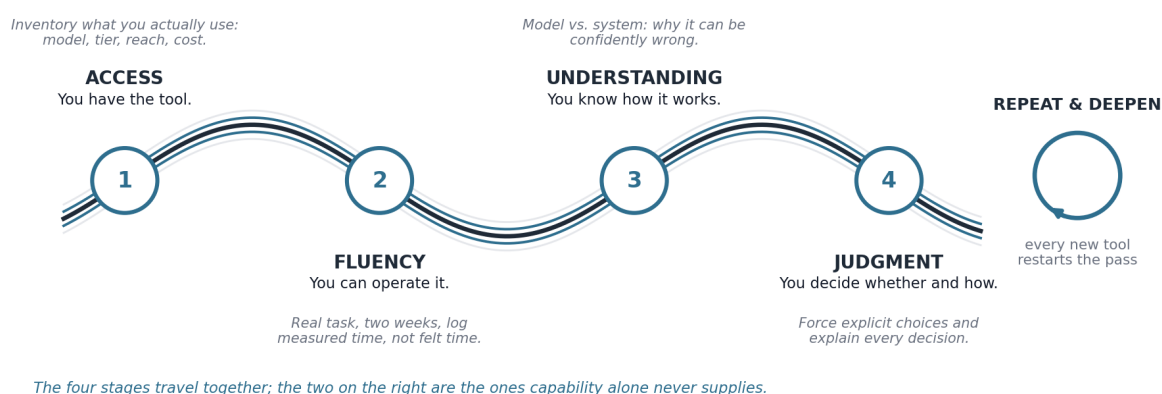
Most confusion about AI comes from collapsing distinct things into one. We have mistaken access for competence, and competence for literacy: having a chatbot is not knowing how to use it, knowing how to use it is not understanding it, and understanding it is not a matter of more commands but of judgment. Natili therefore keeps two axes apart. Literacy describes what you know: *Access* (you have the tool), *Fluency* (you can operate it), *Understanding* (you know enough about how it works to evaluate what it gives you), and *Judgment* (you can decide when, where, how, and whether it should be used at all). Capability describes what you can do: *Using* (direct prompting for everyday tasks), *Mastering* (running a multi-step process across several tools with your own judgment holding it together), *Building* (creating skills, agents, and agentic configurations), and *Architecting* (designing the full system that lets a model serve real users reliably). Almost everything the industry teaches sits at the hinge, *Fluency*. From there, one direction is more capability, which is what vendors sell; the other is *Understanding* and then *Judgment*, which is what keeps its

value when the tools change. *Literacy* comes first because it determines which capabilities are worth having and how to use them.

Natili's Applied AI Literacy framework is informed by, but does not restate, the AI Fluency Framework of Rick Dakan and Joseph Feller,<sup>22</sup> whose four connected competencies (delegation, description, discernment, diligence) describe mature human-AI collaboration as one set. The author separates two dimensions that are usually taught and measured together and uses fluency in a deliberately narrower sense: the operational ability to give useful instructions and get useful results. On that definition, description corresponds to the paper's Fluency stage, and the other three competencies extend into Understanding and Judgment. That is not a disagreement with the 4D framework but a choice of analytical emphasis: unbundling maturity into two axes so that a learner, educator, or organization can tell the difference between becoming more capable with AI and becoming better able to evaluate, govern, and refuse its use. That separation is the paper's original contribution; the two progressions are analytical frameworks, not validated scales or maturity models.

## THE FOUR STAGES OF LITERACY

what you know about these systems · not a destination



**Access:** Having the tool is more than having a login: it means knowing which model answers, which tier you are on, what the tool can reach, and what it costs in money and data. Adoption has outrun understanding: in early 2026, about half of U.S. adults reported using AI chatbots, up from a third two years earlier, with looking things up the most common use.<sup>23</sup> The first literate act is an inventory of what you are actually using.

**Fluency:** Operating the tool effectively, the point where literacy and capability coincide. The gains are real and largest for the least experienced, which is why the industry stops here: in a preregistered experiment with 453 professionals, ChatGPT cut writing time by 40% and raised graded quality by 18%;<sup>24</sup> support agents resolved

14% more issues per hour, with 30–35% gains for novices.<sup>25</sup> Those gains apply to tasks the tool handles well; they say nothing about the tasks it handles badly.

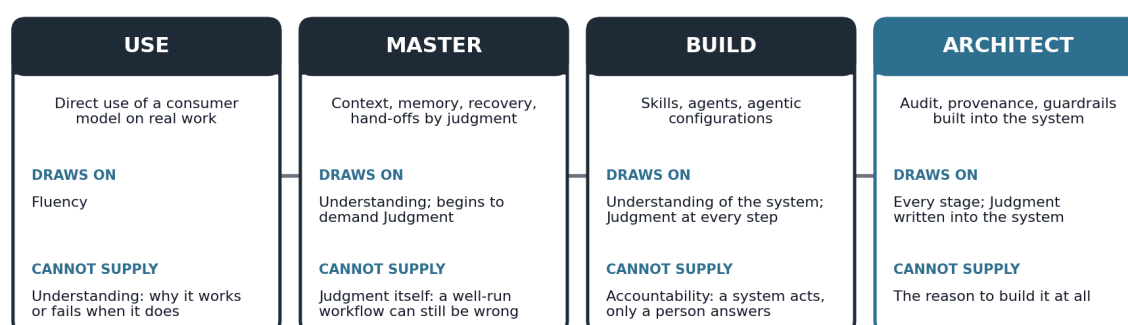
*Understanding:* Knowing why a system behaves as it does and where its limits begin, without needing computer-science expertise: what a transformer does when it predicts the next token, what training does and does not mean, and the difference between the model and the system around it (retrieval, tools, orchestration, guardrails). Confident mistakes are not a defect awaiting a patch; training and evaluation reward guessing over admitting uncertainty, so a model that cannot tell a false statement from a true one states the false one fluently.<sup>26</sup> Understanding is the only defense against that, against the uneven edge of competence, and against over-trusting fluent output, because each is invisible from the interface.

*Judgment:* Deciding whether, when, and how AI should be used, how much confidence its output deserves, and where human responsibility must remain. It begins with the obligation to assume accountability for the output: a system can generate, recommend, and act, but it cannot bear responsibility. That is the core of Digital Dignity, the principle that a person keeps authority over, and accountability for, decisions that affect people even when a system generated the work. Natili uses the term in a narrower operational sense than its wider use in scholarship and policy: as a normative principle for AI governance, not a full account of dignity in digital environments.<sup>27</sup> Ethics guides the choice; laws and policies constrain it but do not replace it. The test of Judgment is whether you can explain the decision: which boundary you chose, what evidence supports it, and which values or consequences shaped it. It develops through decisions, not tool use: should AI be used here at all, what may it do, what must be verified, who is affected, who remains accountable.

*From knowing to doing: four capabilities, read through literacy*

## FROM KNOWING TO DOING

each capability draws on a stage of literacy and exposes the one it lacks



Ordered by what they ask of you, not by rank. Enter at the stage you have reached.

*Using the tools.* Direct use of a consumer model on work you already know. The defensible evidence is more modest than the marketing. Consultants using GPT-4

completed tasks faster and better inside the model's competence, but on a task outside its "jagged frontier" they were 19 percentage points less likely to reach the right answer, because the model sounds equally sure either way.<sup>28</sup> People also misjudge how much they are helped: across three preregistered studies, they overestimated the time AI saved on simple tasks and kept using it when it saved nothing, the "efficiency-gain illusion".<sup>29</sup> Feeling helped is not evidence of being helped; the first literate habit is to record measured time, not felt time. This capability asks only Fluency and cannot supply Understanding.

*Mastering the tools.* Managing context, memory across sessions, failure recovery, and work carried by hand across chat, code, search, and files, with your judgment as the connective tissue. That role is demanding because attention does not parallelize: starting agents is easy, integrating their output is the real work, the "orchestration tax."<sup>30</sup> Mastery also includes right-sizing, matching the model to the task instead of defaulting to the frontier tier. It draws on Understanding and begins to demand Judgment; it cannot supply the Judgment itself, since a well-run workflow can still be the wrong workflow.

*Building with the tools.* Creating three distinct things: a skill (scoped, explicitly invoked), an agent (semi-autonomous, holding a goal and state), and an agentic configuration (several of each, organized toward a larger goal). What matters is the level of autonomy and the handling of state, because they determine the failure modes and the blast radius. The large numbers appear here and must be read carefully: cost-per-task reductions on repeatable work are not a claim that any individual becomes many times more productive, and in one randomized study experienced developers using AI tools took 19% longer while believing they were faster.<sup>31</sup> *Building demands Judgment at every step*, what the agent may do alone and where a human must remain, and cannot supply accountability: a configuration can act, only a person can answer for the action.

*Architecting with the tools.* Turning a model into a system people can rely on: retrieval, decomposition, preserved context, guardrails, self-checks, and visible failures. This is where audit enters, the demand that a system account for itself, and where provenance, the record of where each piece of information came from, becomes the backbone of accountability. Retrieval supplies a provenance trail, evaluation records come from model-as-judge checks, and an immutable ethics policy layer acts as a runtime guardian. Ethics shows up here as design, not commentary. *Architecting is Judgment written into the system*; it presupposes every stage of literacy and cannot supply the reason to build the system at all.

Applied AI Literacy means different things for our studied populations:

- Educators should teach the stages, not the products, and put Understanding and Judgment on the syllabus beside prompting, because a course that stops at Fluency teaches the part that expires first;
- Small businesses should define the problem before choosing the model and choose the smallest tool that does the job;
- Every reader (and/ or consumer) should know which stage they have reached and take the next one on purpose;

The full paper develops further concepts worth the read, among them *meta-prompting*, *right-sizing*, the *distinction between a model and the system* around it, and *auditability by design*.<sup>21</sup>

Thriving in the age of AI is possible only if AI literacy is implemented well and clearly equips people and organizations with knowledge and understanding of AI and robotic systems. Today's instruments that would measure the impact were built for students, not for the further groups requiring knowledge and understanding of AI.<sup>7, 16, 17</sup> Section's June 2026 AI Proficiency Report found that 82 percent of US knowledge workers could not identify what an AI agent is and fewer than one in ten could define one in their own words, at the moment their employers were deploying agents.<sup>32</sup> Cedefop's probability-based survey across eleven EU member states assessed 40 to 60 percent of workers as having a poor understanding of AI technology and its implications.<sup>33</sup> Our own surveys, reported in Chapter 3, found the same pattern in every population: between 86 and 97 percent use AI tools, while only 27 percent of students recognize the concept of AI bias, 7 percent of teachers know what an open-weight model is, and the most requested training topic among SMEs, chosen by 60 percent, is "what is AI?" These are not populations that lack access to training. They are populations that have been trained on tools and left ignorant about the technology.

### ***AI Literacy Content Is the Milestone - in Five Blocks***

If AI literacy is to mean anything, its content must be specified, and it must be the same content whether the learner is a student, a teacher, an experienced professional, or a business owner, adjusted in depth but not in substance. The AI Literacy Study proposes five blocks. None of them requires a computer science degree.

*First, what AI and genAI are and how they differ from biological intelligence.* A learner must be able to explain, in plain language, that a large language model is a next-token prediction architecture grown from data rather than built from rules; that it cannot learn after training, cannot model other minds, and has no truth-seeking mechanism; and that what vendors call a hallucination is therefore a feature of the mathematics, not a bug in the release. This block, set out in Chapter 2, is where almost every misplaced expectation for AI begins, and it is the block that fluency programs skip entirely.

*Second, the realistic risks.* Structural bias that cannot be filtered out of internet-scale data; malicious use; calibration and overreliance, including the cognitive debt that every one of our four populations named in its own words; and catastrophic risk treated as an engineering discipline rather than a horror story. A learner should be able to distinguish the risks that are permanent features to be mitigated from those that are crimes to be prosecuted or design failures to be corrected. Besides, an AI literate individual would be able to critically assess claims from AI labs and BigTech companies when it comes to AI capabilities and risks related to them.

In September 2026, Anthropic's chief executive Dario Amodei published a 3,800-word essay arguing that “we must slow the pace at which we improve the capabilities of AI models,”<sup>34</sup> warning of scenarios in which autonomous AI swarms could hijack global internet infrastructure within the next year. OpenAI's Sam Altman publicly agreed that the industry must “pace the frontier,”<sup>35</sup> while Elon Musk endorsed the call.<sup>36</sup> *These are striking statements, and they deserve to be taken seriously. But they also deserve to be read against the balance sheets of the companies making them.*

Those balance sheets tell a sobering story. Audited financial statements for 2025 - leaked to blogger Ed Zitron and independently verified by the Financial Times, though never officially released by the company - show that *OpenAI* recorded an operating loss of \$20.9 billion on revenue of \$13.1 billion: *the company lost more from its operations than it earned from all of its customers combined.*<sup>37</sup> Analysts at HSBC Global Investment Research calculated in November 2025 that, even under generous revenue assumptions, OpenAI would need roughly \$207 billion in new financing by 2030 to meet its obligations (obligations that include compute and infrastructure commitments with Oracle, Microsoft, Amazon, and Nvidia collectively estimated at \$1.4 trillion, a figure OpenAI's own leadership has publicly acknowledged.)<sup>38</sup> The forward picture is no more reassuring. According to internal projections reported by The Information, *OpenAI expected in September 2025 to burn \$115 billion in cash through 2029; by February 2026 it had revised that forecast upward by a further \$111 billion through 2030, with total spending on training and running its models projected at \$665 billion, and no positive cash flow expected before 2030.*<sup>39</sup> These commitments are contractual pledges rather than cash already spent, and the projections are forecasts rather than audited results; but the direction of every revision to date has been toward more spending, not less. Nor is this pattern limited to OpenAI: Anthropic, the lab most explicitly self-branded around safety and ethics, agreed to \$517 billion in compute commitments covering 14.8 gigawatts of capacity in the eleven months through August 2026, according to The Information<sup>40</sup> - even as its chief executive publicly urged the industry to slow down in September 2026.

In short, the authors of this AI Literacy Study believe it is *unlikely that any plausible stream of future profits will repay capital deployed at this scale.* In that light, *existential rhetoric performs useful commercial work: it justifies valuations that*

*ordinary software economics cannot, positions incumbents as the responsible stewards regulators should entrench, and reframes staggering losses as the necessary price of managing a civilizational transition rather than evidence of a business that does not work as the rest of economy.* It is also worth remembering that these AI CEOs are not disinterested observers of the risk they describe: they are billionaires whose fortunes, reputations, and legacies are bound up in the story that what they are building is world-historically important.

To be clear about our own position: in Chapter 2 we offered a constructive view on mitigating catastrophic risks from AI, and nothing here should be read as a claim that such risks do not exist, or that AI development should continue at its current speed. Some of the concerns raised by these executives are shared by independent researchers with no equity at stake and dismissing them wholesale would be as lazy as accepting them uncritically. Our argument is narrower and, we think, more useful: when the people issuing the gravest warnings are the same people raising hundreds of billions of dollars on the strength of the technology they warn about - and delaying the very IPOs that would expose their finances to public scrutiny - their claims should be examined through the prism of their commercial and personal interests.

*Third, the cost drivers.* Model pricing and how tokens are metered; hardware choices and what they mean for large and small organizations; data and technology sovereignty; the location of the organization and its compliance regime; and literacy itself, because every other cost item can only be assessed by people who understand the technology. Chapter 2 described and explained these. A decision-maker who cannot explain why the visible license fee is the smallest part of an AI implementation is not AI-literate, even if he or she is fluent in deploying and using AI tools.

*Fourth, governance from strategic vision, with clear vocabulary.* AI governance is always secondary to strategic clarity. A learner should know how to run a needs analysis - the four boxes of known and unknown, met and unmet needs described in Chapter 2 - before choosing a tool; should know the sequence of what kind of AI, where in the value chain, how to source it, and only then at what cost; and should understand that a compliance checklist is not governance. This block applies unchanged to a school director deciding on a tutoring platform, an SME owner weighing an agent framework, and a minister drafting a national strategy. Vocabulary used to describe AI is important. Words such as “reasoning,” “understanding,” “agent,” and “hallucination” redistribute responsibility as quietly as they describe capability. A literate person can say which of these terms is a metaphor and which is a mechanism and can insist on the mechanism where consequences follow - in a procurement specification, a safety case, or a statute.

*Fifth, AI Literacy in context - supported by (always updated) data.* AI does not arrive in the world in general; it arrives in a hospital ward, a claims department of an insurance company, a classroom, a ministry, a food bank. The domain determines

how underlying data is treated (collected, stored, pre-processed, used for training and building of AI models, and adjusted), what the technology does, what can go wrong, and who bears the consequences. The practical conclusion for any institution - a company, a school, a government agency, an NGO - is the same: begin not with AI but with your own work and/or field of expertise. Inventory the systems actually in use, name the decisions they touch and the people they affect, and only then teach - the concepts through your cases, the risks through your incidents, the ethics through your dilemmas. Domain knowledge can dictate whether your AI implementations succeed or fail; the same is true of literacy. An AI Literacy course must offer frameworks allowing organizations to diagnose their domain.

The AI Literacy study relied on original surveys designed to capture what different populations actually know and understand about AI and AI tools - and, just as importantly, what they want to know. *Rather than inferring literacy levels from proxies such as usage statistics or self-reported confidence alone, we collected primary data directly from respondents* and mirrored the full results in Chapter 3. Going forward, edutAIment will *extend this data collection per industry and geography*, and in some cases will narrow the analysis to specific age groups where finer resolution yields better clarity. This data matters because it provides a baseline: *a decision maker can use it to design AI literacy initiatives tailored to the context of their own business environment, their educational institution, their NGO, or their governmental department - grounding those programs in evidence about what their specific population knows, misunderstands, and wants to learn*, rather than in assumptions borrowed from other settings.

Fluency - how to use a given tool - sits on top of these five blocks. Taught in this order, fluency becomes portable: a learner who understands why models fail can evaluate any model and will not need retraining when the vendor changes. *The gains from AI are real, uneven, and bounded, and they accrue to organizations that invest in workflow design, evaluation, and governance rather than to those that simply adopt the newest model.*<sup>41</sup> The evidence Natili reviews points the same way - benefits that reverse beyond the “jagged frontier,” people who overestimate the time AI saves, experienced developers who work slower while believing they are faster.<sup>28, 29, 31</sup> The gap between felt and measured performance is precisely the calibration failure that literacy must close.

### ***Principles to Implement AI Literacy***

A novel approach to AI literacy should rest on four principles. They are not abstract commitments but design choices, each with direct consequences for what is taught, in what language, and to whom.

*The first principle is human agency.* The fundamental belief of this study, stated in its abstract and earned across its chapters, is that people - not models, not platforms, not frameworks - decide whether AI makes societies stronger or more fragile. A

literacy program anchored in human agency therefore measures its success differently from a fluency program. The question is never how many people used the tool, how many licenses were activated, or how many certificates were issued. The question is whether, after the program, a person can do something they could not do before: explain a model's failure to a colleague, challenge a vendor's claim, refuse a deployment that solves no defined need, or design a lesson, a workflow, or a policy in which AI amplifies internalized human capability instead of replacing its formation. Agency is the output. Everything else is instrumentation.

*The second principle is interdisciplinarity.* Supporting fundamental research in developing new AI technologies is paramount. This study has argued, from Chapter 2 onward, that several of AI's gravest risks are structural: bias and opacity are properties of the current mathematics, and no compliance regime can remove what the architecture guarantees. The way out is not less technology but better technology exploring causality, world models, and theory of mind. Literacy without a STEM pipeline consumes the current technology forever; a STEM pipeline without literacy builds technology no society is prepared to govern.

Chapter 2 showed how the vocabulary problem needs linguists who can say what a term presupposes, engineers who know which technical claim it stands in for, and legal scholars who know what it triggers in a statute. The risk picture needs mathematicians and psychologists in the same room as auditors and educators. The cost picture needs economists and experts in policy making.

Creative industries shape our perception of AI - and artists are a seismograph for what might happen. Most people's picture of artificial intelligence was formed by fiction, film, and art long before any classroom or framework reached them, and that picture governs how they vote, buy, fear, and adopt. A literacy program that ignores the creative industries cedes the most powerful channel of public understanding to whoever fills it - today, largely to marketing on one side and dystopia on the other. But the relationship runs deeper than communication. Artists register the tremors of a technology before its earthquakes arrive: they were the first to experience, in their own labor and livelihoods, what training data appropriation means, what synthetic content does to authorship, what it is to be imitated at scale by a system that never asked. The disputes over creative rights and training data are not a side issue to AI literacy; they are one of its clearest live lessons in provenance, permissible use, and the economics of the technology. Furthermore, entertainment industry enables a better communication of AI-related content to different age groups.

*The third principle is using simple language* while designing and introducing AI literacy programs. The evidence from science communication research is unambiguous: jargon does not merely bore audiences, it actively damages understanding and distorts risk perception. In a landmark experiment published in 2019, researchers Bullock, Amill, Shulman and Dixon found that the presence of jargon impairs people's ability to process information about emerging technologies,

and that this impairment triggers resistance to persuasion, increases risk perceptions, and lower support for adopting the technology.<sup>42</sup> Shulman and colleagues showed that jargon-laden messages also increase skepticism and reduce the perceived credibility of the source.<sup>43</sup> Applied to AI, the implication is uncomfortable for experts: every time we explain "transformer architectures" or "reinforcement learning from human feedback" to a lay audience without translation, we are not educating them - we are simultaneously confusing them and making them either dismissive or disproportionately afraid. Researchers working on public-facing AI risk communication have reached the same conclusion from the opposite direction. A 2025 study behind the "Atlas of AI Risks" found that existing AI risk resources, built by and for experts, are so technical that ordinary people cannot connect them to their own lives<sup>44</sup> - even though it is precisely this understanding that enables citizens to make informed decisions, advocate for their interests, and push for better regulation.

This is why the democratization of AI knowledge cannot be treated as a work of charity. It is the mechanism by which societies become capable of governing the technology at all. The best-documented proof that it works is Elements of AI, the free online course created in 2018 by the University of Helsinki and MinnaLearn as part of Finland's national ambition to teach AI basics to one percent of its population. Deliberately designed to require no mathematics, no coding, and no prior experience, the course - whose reach and unusually diverse audience were described earlier in this chapter - was offered by the Finnish government as a gift to all EU member states during Finland's 2019 EU Council presidency, the same year it won MIT's Inclusive Innovation Challenge and was ranked the world's best online computer science course by Class Central. Elements of AI demonstrated that when AI is explained in ordinary language, the appetite for understanding it is nearly universal.

A growing ecosystem of communicators has since proven that clarity and rigor are not opposites. Ethan Mollick, a Wharton professor, reaches hundreds of thousands of readers through his newsletter One Useful Thing<sup>45</sup> and his bestselling book "Co-Intelligence"<sup>46</sup> by writing about AI capabilities and risks in the language of everyday work rather than computer science. Grant Sanderson's 3Blue1Brown channel has shown that even the internal mechanics of neural networks and large language models can be made visually intuitive for a mass YouTube audience,<sup>47</sup> while Andrej Karpathy - a founding member of OpenAI and former head of AI at Tesla - has published freely available lectures that build a language model from scratch in front of the viewer.<sup>48</sup> Andrew Ng's non-technical course AI for Everyone (DeepLearning.AI)<sup>49</sup> and the explanatory journalism of outlets such as MIT Technology Review explain AI in simple language and reach millions of people. What unites these providers is not simplification for its own sake but respect: they treat the public as capable of genuine understanding, provided someone does the

work of removing the linguistic barriers. That is the standard this study argues AI communication - and especially AI risk communication - must meet.

edutainment's own contribution to this landscape follows the same conviction. "AI Snacks with Romy & Roby," the podcast by Anastassia Lauterbach, translates AI and robotics from complex scientific concepts into plain-language conversations accessible to teens, parents, teachers, and anyone curious about how these technologies actually work. The podcast is deliberately positioned as "The Science Behind" companion to "Romy, Roby and the Secrets of Sleep," the first "AI adventure for the whole family": a narrative in which the story carries the reader and a scientific appendix explains the AI, robotics, and quantum computing concepts woven into the plot. Its strengths are real: the edutainment structure - narrative first, science appendix second, podcast as the living bridge between them - is genuinely rare in the AI literacy landscape. The limitations are equally real: unlike Elements of AI, with its two million learners and government backing, Romy & Roby is an independent project whose reach currently stands at roughly 25,000 listeners and readers, and no independent evaluation has yet measured what its audience actually retained from learning.

*The fourth principle is about customized method of delivering AI literacy content.* The five content blocks might be universal, even if the extent of the material varies with the age of a learner, his or her level of curiosity, and willingness to be exposed to additional information sources. The method of delivering AI Literacy content is however not homogeneous, because the populations of Chapter 3 face different binding constraints. A method that works for a mid-career professional with a subscription and a free weekend fails a school with no budget and a teacher shortage, and fails again for an SME owner whose scarcest resource is decision-making time. The program therefore must pair the common content with population-specific methods.

Building on these four principles, the following sections introduce further practical suggestions for implementing AI literacy programs in SMEs and schools.

### ***SMEs: Benchmarking as one of the tools to boost AI implementations***

Small and mid-sized businesses are surrounded by thousands of AI tools, agents, and platforms, marketed with claims no buyer can verify, priced in ways that conceal the true cost of adoption, and churning faster than any procurement cycle. The result is documented in our survey: *near-universal usage, strategic confusion, shadow AI tolerated by default, cost forecasts missed by half, and - from businesses already running agentic systems - the plea to be told, at last, what AI actually is.* An SME cannot out-spend an enterprise on evaluation. It cannot run three parallel pilots, hire a machine-learning team, or absorb a failed implementation. *What it can do is out-*

*decide an enterprise* - if someone gives it the decision strategic direction and infrastructure.

That *infrastructure is independent benchmarking*, and the program treats it as a core literacy method for this population. Any benchmarking activity on existing AI tools is beneficial to SMEs, because the binding problem is that they cannot make sense of what is available and what fits. This AI Literacy Study therefore calls for systematic, independent, regularly updated benchmarks of AI tools by business function and industry. Regular updating is not a luxury: a 2026 analysis accompanying the latest wave of evaluation funding found twenty-nine of sixty widely used public benchmarks already saturated, their tasks effectively memorized by the models they were meant to test.<sup>50</sup>

The question of *who benchmarks AI tools* - not “just” the underlying language models, but the workflow platforms, agents, insight engines, and generative applications built on top of them - reveals *an ecosystem that is growing explosively, yet remains fragmented into layers that answer different questions, and leaves a conspicuous gap precisely where buyers need evidence most*.

The newest and most dynamic layer consists of independent commercial benchmarkers, and the capital flowing into them confirms that such a market meets demand.

Vals AI, a San Francisco firm founded by Stanford-trained engineers, evaluates AI systems on real professional tasks - legal research, financial analysis, tax questions, document-heavy workflows, and multi-step agent tasks - using privately held test suites that cannot leak into training data.<sup>51</sup> In August 2026 it raised a \$40 million Series A led by Andreessen Horowitz at a reported \$400 million valuation.<sup>52</sup> Its Vals Legal AI Report broke genuinely new ground: a first-of-its-kind study evaluating how commercial legal AI products perform across seven legal tasks, benchmarked against a control group of practicing lawyers. This is the rare example of what buyers actually need - *named products measured against human professionals on real work*.

Academic evaluation confirms how wide the gap between marketing and capability remains: in Carnegie Mellon's TheAgentCompany, a simulated software company in which agents must complete a full day of knowledge work, the best-performing agent finished only about thirty percent of its tasks.<sup>53</sup> LMArena, whose leaderboards draw more than five million users a month, saw its valuation rise from \$600 million in May 2025 to \$1.7 billion seven months later, on roughly \$30 million in annualized revenue.<sup>54</sup> Artificial Analysis publishes cross-model leaderboards with transparent methodology and is expanding into applied enterprise benchmarks covering analyst agents, IT operations, and legal workflows.<sup>55</sup> Evaluation infrastructure vendors such as LangSmith, Galileo, DeepEval, and Comet Opik which sell enterprises the tooling to benchmark their own agents rather than publishing public rankings – aren't a good fit for SMEs due to their fee structures. edutAlnment won't recommend analyses by

the established analyst firms, such as Gartner's Magic Quadrant, Forrester's Wave, or Everest Group's PEAK Matrix. Although they today cover agentic AI products, conversational AI platforms, customer service, employee services, and enterprise search, their insights are built on vendor briefings, product demonstrations, and selective customer references - not empirical benchmarks with test suites and task-completion scores. We believe that further work - preferably at the local level, at chambers of commerce, industrial associations, and local universities must happen to help local companies in their AI implementation efforts – while benchmarking AI providers against the real needs of their local economic ecosystem.

### ***Schools: AI as Infrastructure for Personalized Learning***

For schools, the program draws a line that current policy refuses to draw: rolling out general-purpose LLMs across classrooms will not be beneficial. Chapter 3 documented both sides of this line. On one side, chatbots issued to students before anyone defined what problem they solve, and the cognitive-debt evidence - from the MIT Media Lab work<sup>56</sup> to Brookings' fifty-country study, both reviewed in Chapter 3 - accumulating in exactly the population least able to afford it. On the other side, South Korea's reversal within a semester shows what happens when technology is pushed into schools faster than the teachers who must carry it. Forbidding AI and blindly signing LLM licenses for a campus are two extremes, and both damage students' chances to thrive.

Between the extremes sits the application the program endorses: *AI that personalizes learning*. This is AI underneath and humans on top, the architecture the Alpha School case in Chapter 3 illustrated, and the direction of Michael Leggs' essay tool, which coaches a student through writing rather than writing for them. Used this way, AI does not replace the formation of a mind; it locates where formation is failing and concentrates human and machine effort there.

The EdTech market suffers the same opacity as the SME tool market, with higher stakes: the buyers are school directors and ministries, the “subjects” are children, and the marketing is indistinguishable from the evidence. The program therefore calls for independent benchmarking of personalized-learning tools on the dimensions that matter - demonstrated learning efficacy measured against cognition rather than logins, as Estonia's AI Leap measures it; pedagogical design, including whether the tool builds capability or dependence; data protection appropriate to minors; total cost for a public-school budget; and teacher experience, with teachers inside the evaluation loop rather than downstream of the purchase. And it calls for public and philanthropic funding programs that scale what the benchmarks validate: money should flow to the tools that prove effective across the widest range of schools. Independent creators should be helped and celebrated, not left within the borders of their individual social networks and geographies.

## A Market for Trust in Intelligence?

As big and small companies operate alongside smart machines (with consumers buying their products and services, and therefore being impacted by their decisions), *the verification of AI tools may well become a market in its own right - an industry whose product is not intelligence but trust in intelligence*: independent attestation that a tool's underlying data meets defined standards of accuracy, completeness, timeliness, and consistency, that its sources are real and its provenance is declared. This is a call going beyond benchmarking of commercial tools, and its realization requires advanced level of AI literacy.

There is an instructive precedent: when the online world emerged in the late 1990s, websites began displaying trust badges - TRUSTe, BBBOnline, VeriSign, WebTrust, eKomi - issued by organizations claiming to have checked a site's privacy and security practices. These seals were themselves a commercial product, sold to businesses for annual fees, and their credibility eroded accordingly: researchers found some certified sites were no more trustworthy than uncertified ones, regulators fined seal providers for skipping the very checks they promised, and some badges required nothing but payment.

*The lesson for an emerging AI-verification market is that the verifiers must themselves be verifiable - through independence from those they certify, transparent methodology, and consequences for failed audits - because a badge whose business model is selling reassurance measures reputation management, not quality, and simply relocates the trust problem rather than solving it.*

### Decomposing “AI Solution Quality”

An AI tool is only as trustworthy as the data underneath it, and “quality” is too broad a word to be useful. The same is true of the fashionable demand for “AI-ready” data: nobody can act on it. *This study proposes double-clicking on quality and breaking it into empirically testable components - a framework summarized as ACT+C: Accuracy, Completeness, Timeliness, and (cross-border) Consistency.* The labels are interchangeable (credibility for accuracy, latency for timeliness); the discipline of decomposing quality is not. Anthony Scriffignano, former Chief Data Scientist of Dun&Bradstreet, always applied this framework while advising product teams within the company, enterprise clients, US government and international regulators.<sup>57</sup>

*Accuracy:* Accuracy means different things for different data. A claim can be accurate because it is confirmable against an authoritative source, or because it comes from the authoritative source itself. Much of what feeds AI systems fits neither bucket - it is modeled or projected.

Accuracy also has both a colloquial and a mathematical meaning (performance against a known test set), and the mathematical one breaks down in highly dynamic systems: if the data about a thing changes more slowly than the thing itself, you are

learning how it was, not how it is. Accuracy can therefore only ever be partially measured, except for dispositive facts about the past.

*Completeness:* Not everything a decision requires is on the internet or in the database. *The honest question is not “do I have all the data?” (you never do) but the elasticity question: how wrong can I be and still responsibly make the decision I am making?* Knowing you lack enough information and deciding anyway is negligence. Before ingesting a new source, ask whether it will make the system smarter or just add noise.

*Timeliness:* *Data collected yesterday is not simply “one day old” - its value depends on how fast the system being studied is changing relative to the age of the data.* Availability itself shifts: documents that are public one day can be sealed the next, so an AI trained today can honestly be “less informed” today than yesterday. Latency relative to the rate of change of the world is the real measure.

*Consistency:* True data is not always the whole truth: datasets carry sub-modes where the same field means different things (just think about whether “the number of employees” at different businesses mean the same thing). Understanding consistency requires separating random-cause from assignable-cause variation - without that study, the honest answer is “you don't know.” Consistency is proved through triangulation: reaching the same conclusion through two or more mutually uninformed routes.

### *Provenance and the Database Behind the Tool*

*Trust also requires transparency about representation: is a claim first-party (experienced directly), second-party (synthesized from trusted sources), or third-party (a trusted source citing another)? All three are legitimate - if declared.*

The same discipline applies to the database behind any AI tool or benchmarking product: know the population, whether the sample is representative and stratified, and what the data can and cannot say (e.g., “This first AI Literacy Study says nothing about teachers in Ireland - I wasn't studying them explicitly”).

A useful practical test of a tool that every individual using an LLM can apply: Ask it for sources and check all of them. Wrong authors, mismatched topics, and dead references are direct evidence about the data layer.

### *AI Literacy and the Accountable Steward*

Applying the ACT+C framework is not a casual exercise - it demands an advanced level of AI literacy. Asking whether a tool's accuracy can be confirmed against an authoritative source, whether its data is complete enough for the decision at hand, how fast the underlying reality is changing relative to the data's age, and whether conclusions survive triangulation across independent sources requires fluency in concepts like representative sampling, random versus assignable cause variation,

and provenance - knowledge that goes well beyond knowing how to prompt a chatbot.

This, in turn, has an organizational consequence: interrogating data quality cannot be delegated to everyone in general and therefore to no one in particular. Every organization - a business, a school, a public agency - needs at least one person who genuinely possesses this expertise: someone who can double-click on the word "quality" when a vendor uses it, ask the Socratic questions ("How do you know the data is right? How wrong can we afford to be and still make this decision?") and translate the answers for colleagues who neither can nor should think this hard about data.

Just as organizations learned that security and privacy require designated, qualified owners rather than diffuse goodwill, trustworthy use of AI requires an accountable steward of data (and AI) quality - and building that role is itself one of the most concrete outcomes an investment in AI literacy can deliver.

### **AI Literacy as a Leadership Decision**

The conventional model of delivery of AI literacy is top-down: a ministry adopts a framework, a curriculum is written, teachers are trained, students are taught, and a decade later the labor market receives literate workers. Chapter 3 showed what happens along that chain. Federal systems cannot agree on a curriculum. Centralized systems that move fast without carrying their teachers, as South Korea did, reverse within a semester. The best-designed national programs, Estonia's and Singapore's, are small-state achievements that took years to build and are not easily copied.

An AI Literacy program - mandated by the law or without such a mandate - might therefore travel a second route, alongside and not instead of national policy. AI literacy is a decision, and the decision can be taken at any scale. The evidence for this claim runs through this entire AI Literacy study.

*One school.* Michael Leggs built an AI that coaches students through an essay rather than writing it for them, in one school, without a mandate. A class of sixth graders in Maastricht, given three one-hour sessions and an adult willing to answer, could explain why current language models are power-hungry and why a model produced historically inaccurate images.

*One community.* Elisa Torres Durney, at sixteen, with one scholarship course and no budget, built Girls in Quantum into a platform reaching more than 8,000 students in 60 countries. A parents' association, a library, or a youth club can run the smaller version of the same play.

*One company.* The SME roadmap in Chapter 3 begins with a one-page inventory, an amnesty survey on shadow AI, and a needs analysis. None of these costs more than a few days of a leader's attention. The case of the Chief Learning Officer who

redesigned training around internalized knowledge, and the retiring-workforce case in which AI was used to capture expertise before it walked out of the door, were decisions taken inside single firms.

*One city, one region.* Chapter 2 described the ecosystems that already exist around most organizations: universities, incubators, data cooperatives, apprenticeship programs, European Digital Innovation Hubs. SMEs collaborating with universities are up to six times more likely to adopt AI. A mayor, a chamber of commerce, or a regional development agency can decide to become the table at which schools, firms, and researchers agree on a shared AI literacy curriculum, a shared benchmark, and a shared talent pipeline.

*Bottom-up delivery has three advantages that top-down delivery cannot match.* It starts now. It adapts content to local language, industry, and need, which no international framework can do. And it produces evidence that top-down policy can later scale. Its disadvantage is the obvious one: it depends on someone deciding.

### ***Recommendations: Literacy First, Capability Read Through It***

The recommendations of this chapter follow one rule, taken from Giorgio Natili's framework and confirmed by every population in this Study: *literacy first, capability read through it*. Literacy is not the first module of an AI program, to be ticked off before the "real" training begins. It is the filter through which every later decision passes, and the precondition for everything that follows.

*For policymakers and standard-setters*

- **Define AI literacy by stages, not by hours or certificates.** Recognize two axes - what people understand and judge, and what they can do with tools - and stop treating fluency in vendor products as the national floor of AI competence.
- **Add the decision tier to role-based frameworks.** Complement tiers such as interface's L0-L2 with a tier for everyone who allocates budget to AI, approves its deployment or sets policy around it. Its evidence is Natili's test of Judgment: a decision that can be explained and defended.
- **Fund measurement before certification.** Commission validated instruments that measure where populations stand on both axes, and support baselines such as the AI Literacy Barometer, extended by industry, geography and age group.
- **Require auditability by design where stakes are high.** Provenance, evaluation records and visible failures make Judgment possible - and make it checkable.

*For organizations, boards and HR*

- **Measure the baseline first.** Know which stage employees, managers and decision-makers have reached on each axis before buying any training.
- **Use literacy as the filter for every AI investment.** No program to build agents or architect systems without the Understanding and Judgment those capabilities presuppose; for every agent in production, name the person who answers for what it does.
- **Go beyond concepts.** Assess literacy where it shows in organizations: decision rights, verification duties, escalation paths, the calibration of human oversight, and the orchestration tax that limits how many agents a person can supervise.
- **Measure time, not feelings.** Record measured productivity against a baseline; felt efficiency is not evidence of efficiency.
- **Appoint accountable stewards of data and AI quality,** able to apply the ACT+C decomposition and to ask vendors how they know their data is right.

#### *For SMEs*

- **Define the problem before choosing the model, and choose the smallest tool that does the job.** Right-sizing is what makes AI sustainable at an SME's price point.
- **Climb the capability axis only as far as judgment reaches.** Most of the near-term value lies in using and mastering tools well; building agents demands explicit limits on their autonomy.
- **Rely on independent benchmarks and one primary vendor in production,** and use local universities, chambers of commerce and industry associations to share the cost of evaluation.

#### *For educators and schools*

- **Teach the stages, not the products.** Put Understanding and Judgment on the syllabus beside prompting; a course that stops at Fluency teaches the part that expires first.
- **Assess through explanation.** Oral examinations and case defenses show Judgment in a way that written output produced with AI cannot.
- **Put AI underneath, humans on top.** Use AI for personalized learning rather than open chatbots, and benchmark EdTech tools against cognition rather than logins.

#### *For individuals*

- **Know which stage you have reached on each axis, and take the next one on purpose.** Start with an honest inventory of what you use (Access), check

what the tool tells you (Understanding), and keep responsibility for what you decide with it (Judgment).

## Sources for Chapter 4

1. Fengchun M, UNESCO et al., *AI Competency Framework for Students*. Aug. 8 2024, Update Jan 16 2026. Available from <https://www.unesco.org/en/articles/ai-competency-framework-students> (cited Sept 11 2026)
2. Fengchun M., UNESCO, et al., *AI Competency Framework for Teachers*. Aug 8 2024, Update Jan 16 2026. Available from <https://www.unesco.org/en/articles/ai-competency-framework-teachers> (cited Sept 11 2026)
3. OECD, European Commission, *Empowering Learners for the Age of AI. An AI Literacy Framework for Primary and Secondary Education*. June 18 2026. Available from [https://ailiteracyframework.org/pdfs/framework\\_pdf/AILF\\_en.pdf](https://ailiteracyframework.org/pdfs/framework_pdf/AILF_en.pdf) ; Rowe H., *Empowering Learners for the Age of AI: Presenting the Finalised AI Literacy Framework for Primary and Secondary Education*. June 18 2026. (both cited Sept 11 2026)
4. Long D. et al., *What is AI Literacy? Competencies and Design Considerations*. ACM (Association for Computing Machinery) Digital Library. Research Article. CHI '20: Proceedings of the 2020 CHI Conference on Human Factor in Computing Systems, Pages 1-16. Apr 23 2020. Available from <https://dl.acm.org/doi/10.1145/3313831.3376727> (cited Sept 11 2026)
5. Ng D. et al., *Conceptualizing AI Literacy: An exploratory Review*. *Computers and Education: Artificial Intelligence*. Volume 2, 2021, 100041. Available from <https://www.sciencedirect.com/science/article/pii/S2666920X21000357> (cited Sept 11 2026)
6. *Regulation (EU) 2024/1689 of the European Parliament and the Council of 13 June 2024 laying down harmonized rules on artificial intelligence*. Document 32024R1689. Available from <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (cited Sept 11 2026)
7. *AI Literacy – Questions & Answers*. Update Sept 16 2026. Available from <https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers> (cited Sept 11 2026)
8. *Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689. (EU 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)*. Document 32026R1755. Available from <https://eur-lex.europa.eu/eli/reg/2026/1744/oj> (cited Sept 11 2026)
9. *Repository of AI Literacy Practices*. Available from <https://digital-strategy.ec.europa.eu/en/policies/repository-ai-literacy-practices> (cited Sept 11 2026)

10. Executive Order 14277 of Apr 23 2025: *Advancing Artificial Intelligence Education for American Youth*. Available from <https://www.federalregister.gov/documents/2025/04/28/2025-07368/advancing-artificial-intelligence-education-for-american-youth>; The White House. 60+ Organizations Sign White House Pledge to Support America's Youth and Invest in AI Education. June 30 2025. Available from <https://www.whitehouse.gov/releases/2025/06/60-organizations-sign-white-house-pledge-to-support-americas-youth-and-invest-in-ai-education/> (all cited Sept 11 2026)
11. Elements of AI. Available from <https://www.elementsofai.com/> (cited Sept 11 2026)
12. University of Helsinki, *Elements of AI Has Introduced One Million People to the Basic of Artificial Intelligence*. May 23 2023. Available from <https://www.helsinki.fi/en/news/artificial-intelligence/elements-ai-has-introduced-one-million-people-basics-artificial-intelligence>; Ministry of Economic Affairs and Employment. Finnish Government. *Elements of AI Course Continues to Increase Artificial Intelligence Skills Among Europeans*. Jan 10 2022. Available from <https://valtioneuvosto.fi/en/-/1410877/elements-of-ai-course-continues-to-increase-artificial-intelligence-skills-among-europeans> (both cited Sept 11 2026)
13. GovUK, Department for science, Innovation and Technology, *Free AI Training for All, as Government and Industry Programme Expands to Provide 10 million Workers with Key AI Skills by 2030*. Jan 28 2026. Available from <https://www.gov.uk/government/news/free-ai-training-for-all-as-government-and-industry-programme-expands-to-provide-10-million-workers-with-key-ai-skills-by-2030> (cited Sept 11 2026)
14. GovUK, *Skills England, AI Foundation Skills for Work Benchmark*, Jan 28 2026. Available from <https://www.gov.uk/government/publications/ai-foundation-skills-for-work/ai-foundation-skills-for-work-benchmark> (cited Sept 11 2026)
15. Eurostat. December 2025. *Use of Artificial Intelligence in Enterprises*. Available from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use\\_of\\_artificial\\_intelligence\\_in\\_enterprises](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises) (cited Sept 11 2026)
16. Weipprecht A., *AI Literacy Under Article 4 of the EU AI Act: Duty, Proof, Training*. Provimedia. Mar 26 2026. Available from <https://www.provimedia.de/en/blog/ai-literacy-certificate-eu-ai-act-2026>; Eu Ai Act. Online Training, Exam, Certificate of Completion. Become an Artificial Intelligence Act Trained Professional (AIActTPro). Available from [https://www.artificial-intelligence-act.com/Artificial\\_Intelligence\\_Act\\_Trained\\_Professional\\_\(AIActTPro\).html](https://www.artificial-intelligence-act.com/Artificial_Intelligence_Act_Trained_Professional_(AIActTPro).html); Coursera. Navigating the EU AI Act: A Playbook for Building an AI-Literate Workforce. Available from

- <https://www.coursera.org/enterprise/resources/ebook/eu-ai-act-playbook> (all cited Sept 11 2026)
17. Velasco L. et al., *An Operational Framework for AI Literacy in the Workplace. Informing Action Plans under the EU AI Act*. Interface. Aug 25 2026. Available from <https://www.interface-eu.org/publications/ai-literacy-workplace-framework> (cited Sept 11 2026)
  18. Lintner T., *A Systemic Review of AI Literacy Scales*. NPJ. Science of Learning. Aug 6 2024. Available from <https://www.nature.com/articles/s41539-024-00264-4> (cited Sept 11 2026)
  19. Cisgrive J., et al., *DigComp 3.0 European Digital Competence Framework*. 2025. Available from [https://publications.jrc.ec.europa.eu/repository/bitstream/JRC144121/JRC144121\\_01.pdf](https://publications.jrc.ec.europa.eu/repository/bitstream/JRC144121/JRC144121_01.pdf) (cited Sept 11 2026)
  20. NIST AI 101-1, *Artificial Intelligence Risk Management Framework*, NIST AI 100-1, January 2023, Available from <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>; ISO/IEC 42001:2023, Available from <https://www.iso.org/standard/42001> (reports costs in Switzerland CHF 225); NIST AI RMF to ISO/IEC FDI 42001 *AI Management System Crosswalk*. Available from [https://airc.nist.gov/docs/NIST\\_AI\\_RMF\\_to\\_ISO\\_IEC\\_42001\\_Crosswalk.pdf](https://airc.nist.gov/docs/NIST_AI_RMF_to_ISO_IEC_42001_Crosswalk.pdf) (all cited Sept 11 2026)
  21. Natili G., *Beyond Fluency*. Sept. 2026. Available from [www.aiedutainment.ai](http://www.aiedutainment.ai) (cited Sept 24 2026)
  22. Dakan R. et al., *AI Fluency Framework: Delegation, Description, Discernment, Diligence*. Ringling College of Art and Design, University College Cork. 2025. Available from <https://aifluencyframework.org/> (cited Sept 24 2026)
  23. Gottfried J. et al., *Americans and AI 2026: Chatbots, Smart Devices and Views on Impact*. Pew Research Center. June 17 2026. Available from <https://www.pewresearch.org/internet/2026/06/17/americans-and-ai-2026-chatbots-smart-devices-and-views-on-impact/> (cited Sept 24 2026)
  24. Noy S. et al., *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*. Science. Vol 381, No 6654, Pages 187-192. July 13 2023. Available from <https://www.science.org/doi/10.1126/science.adh2586> (cited Sept 24 2026)
  25. Brynjolfsson E. et al., *Generative AI at Work*. The Quarterly Journal of Economics. Vol 140, No 2, Pages 889-942. May 2025. Available from <https://academic.oup.com/qje/article/140/2/889/7990658> (cited Sept 24 2026)
  26. Kalai A. T. et al., *Why Language Models Hallucinate*. Sept 4 2025. arXiv:2509.04664. Available from <https://arxiv.org/abs/2509.04664> (cited Sept 24 2026)
  27. Event announcement: Natili G., *The Human Blueprint: Reclaiming Digital Dignity in the Age of Agentic AI*. Mar 19 2026. Available from <https://www.eventbrite.com/e/the-human-blueprint-reclaiming-digital-dignity-in-the-age-of-agentic-ai-tickets-1985205034699> (cited Sept 24 2026)

28. Dell'Acqua F. et al., *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*. Harvard Business School  
[https://www.hbs.edu/ris/Publication%20Files/dell-acqua-et-al-2026-navigating-the-jagged-technological-frontier\\_5c589c8c-fbb5-458f-b285-c944746cd717.pdf](https://www.hbs.edu/ris/Publication%20Files/dell-acqua-et-al-2026-navigating-the-jagged-technological-frontier_5c589c8c-fbb5-458f-b285-c944746cd717.pdf) (cited Sept 26 2026)
29. Yu S. et al., *The Efficiency-Gain Illusion: People Underestimate the Rate of AI Use and Overestimate Its Benefits on Simple Tasks*. May 21 2026.  
arXiv:2605.22687. Available from <https://arxiv.org/abs/2605.22687> (cited Sept 24 2026)
30. Osmani A., *The Orchestration Tax*. May 24 2026. Available from  
<https://addyosmani.com/blog/orchestration-tax/> (cited Sept 24 2026)
31. Becker J. et al., *Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity*. METR. July 2025. arXiv:2507.09089.  
Available from <https://arxiv.org/abs/2507.09089> (cited Sept 24 2026)
32. Section. *The AI Proficiency Report*. June 2026. Available from  
<https://www.sectionai.com/ai/the-ai-proficiency-report> (cited Sept 11 2026)
33. Cedefop, *Skills Empower Workers in the AI Revolution. First Findings from Cedefop's AI Skills Survey*. Available from  
<https://www.cedefop.europa.eu/en/publications/9201> (cited Sept 11 2026)
34. Amodei D., *We Must Pace the Frontier*. Sept. 2026. Available from  
<https://darioamodei.com/post/we-must-pace-the-frontier> (cited Sept 13 2026)
35. Altman S., Twitter/ X @sama, Sept 12 2026, Available from  
[https://x.com/sama/status/2098811563415150910?ref\\_src=twsrc%5Etfw%7Ct\\_wcamp%5Etweetembed%7Ctwtterm%5E2098811563415150910%7Ctwgr%5Eb5a921f4f0a3140aaacafb1003eaf2069977f45f%7Ctwcon%5Es1\\_&ref\\_url=https%3A%2F%2Fiframe.nbcnews.com%2FvTlxpwjc%3F\\_showcaption%3Dtrueapp%3D1](https://x.com/sama/status/2098811563415150910?ref_src=twsrc%5Etfw%7Ct_wcamp%5Etweetembed%7Ctwtterm%5E2098811563415150910%7Ctwgr%5Eb5a921f4f0a3140aaacafb1003eaf2069977f45f%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fiframe.nbcnews.com%2FvTlxpwjc%3F_showcaption%3Dtrueapp%3D1) (cited Sept 12 2026)
36. Sapoot A., *OpenAI Rules Out IPO This Year As Altman, Musk & Amodei Warn AI Is Moving Too Fast*. CNBC, Sept 12 2026. Available from  
<https://www.cnbc.com/2026/09/12/anthropics-amodei-proposes-plan-to-slow-the-pace-of-advancing-ai-capabilities.html> (cited Sept 12 2026)
37. Zitron E., *Exclusive: OpenAI Losses Increased Nearly 8X in 2025, With Spending Hitting \$34 Billion*. June 15 2026. Available from  
<https://www.wheresyoured.at/exclusive-openai-financials/> 2026; verified and reported by the Financial Times, "OpenAI Spending Hit \$34bn Last Year Ahead of Planned IPO, June 16, 2026, available from  
<https://www.ft.com/content/e15b0d7e-ff6b-4f16-ba7a-4068feddb828> (both cited Sept 12 2026)
38. Lichtenberg N., *OpenAI Won't Make Money by 2030 and Still Needs to Come Up With Another \$207 Billion to Power Its Growth Plans, HSBC Estimates*. Nov 26 2025. Available from [https://finance.yahoo.com/news/openai-won-t-money-2030-182441351.html?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2x](https://finance.yahoo.com/news/openai-won-t-money-2030-182441351.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x)

- [LmNvbS8&guce\\_referrer\\_sig=AQAAAE9MXgQT NS3TM1GcoJGStmIUR86v LKfr0YelqpFam22i4LM3JsFw8RPvCiSJO MKIIMj9G0eIVcZxF4HQmsWBI4X 3LIgKGoUGI0nWIOJLviGRabujJFUJ0xj2jbSBBdYXRCDJPbAM00BNWKw6 Fn2sg55WQ1O6wUIT3-1J-aosr](https://www.theinformation.com/articles/openai-says-business-will-burn-115-billion-2029) (cited Sept 12 2026)
39. Muppidi S., *OpenAI Says Its Business Will Burn \$115 billion Through 2029*. The Information. Sept 5 2025, Available from <https://www.theinformation.com/articles/openai-says-business-will-burn-115-billion-2029>; Weinberg C., *OpenAI Projections Imply Losses Tripling to \$14 Billion in 2026*. Oct 9 2024. Available from <https://www.theinformation.com/articles/openai-projections-imply-losses-tripling-to-14-billion-in-2026> (cited Sept 12 2026)
  40. Pau V., *How Anthropic Clinched \$517 Billion in Compute Deals in 11 Months*. Sept 6 2026. Available from <https://www.theinformation.com/articles/anthropic-clinched-517-billion-compute-deals-11-months> (cited Sept 12 2026)
  41. Bain & Company, *Technology Report 2025*. Available from <https://www.bain.com/insights/topics/technology-report/> (cited Sept 11 2026)
  42. Bullock, O. M. et al., *Jargon as a Barrier to Effective Science Communication*, Evidence from Metacognition. Public Understanding of Science. 2019 Oct, 28(7): 845-853. NIH (National Library of Medicine). PubMed. Available from <https://pubmed.ncbi.nlm.nih.gov/31354058/> (cited Sept 12 2026)
  43. Shulman, H. C., et al., *The effects of jargon on processing fluency, Self-Perceptions, and Scientific Engagement*. Journal of Language and Social Psychology. Jan 2020. ResearchGate. Available from [https://www.researchgate.net/publication/338897373\\_The\\_Effects\\_of\\_Jargon\\_on\\_Processing\\_Fluency\\_Self-Perceptions\\_and\\_Scientific\\_Engagement](https://www.researchgate.net/publication/338897373_The_Effects_of_Jargon_on_Processing_Fluency_Self-Perceptions_and_Scientific_Engagement) (cited Sept 12 2026)
  44. Bogucka et al, *Atlas of AI Risks: Enhancing Public understanding of AI Risks*. Feb 7 2025. arXiv:2502.05324 Available from <https://arxiv.org/html/2502.05324v1> (cited Sept 12 2026)
  45. <https://www.oneusefulthing.org/> (cited Sept 12 2026)
  46. Mollick E., *Co-Intelligence. Living and Working with AI*. Penguin. 2024
  47. <https://www.youtube.com/c/3blue1brown> (cited Sept 12 2026)
  48. [https://www.youtube.com/playlist?list=PLAqhlrjxW9U8-vZ\\_s\\_cjKPT\\_FqRStI](https://www.youtube.com/playlist?list=PLAqhlrjxW9U8-vZ_s_cjKPT_FqRStI) (cited Sept 12 2026)
  49. <https://www.deeplearning.ai/courses/ai-for-everyone> (cited Sept 12 2026)
  50. Pebblous, *Vals AI \$400M Valuation and Benchmark Saturation*. Aug 2026. Available from <https://blog.pebblous.ai/blog/vals-ai-eval-as-infrastructure/en/> (cited Sept 14 2026)
  51. <https://www.vals.ai/home> (cited Sept 12 2026)
  52. <https://a16z.com/announcement/investing-in-vals/> (cited Sept 12 2026)
  53. Xu F. F. et al., *TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks*. Carnegie Mellon University. Dec 2024.

arXiv:2412.14161. Available from <https://arxiv.org/abs/2412.14161> (cited Sept 14 2026)

54. <https://arena.ai/>, Bort J., *LMarena Lands \$1.7B Valuation Four Months After Launching its Product*. TechCrunch. Jan 6 2026.. available from <https://techcrunch.com/2026/01/06/lmarena-lands-1-7b-valuation-four-months-after-launching-its-product/>
55. <https://artificialanalysis.ai/> (cited Sept 12 2026)
56. Kosmyna N. et al., *Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for an Essay Writing Task*. MIT Media Lab. June 2025. arXiv:2506.08872. Available from <https://arxiv.org/abs/2506.08872> (cited Sept 14 2026)
57. Similar decompositions of data and AI trustworthiness appear in: Wang et al., *Beyond Accuracy: What Data Quality Means to Data Consumers*, Journal of Management, Mar 1 1996, Available from <https://www.semanticscholar.org/paper/Beyond-Accuracy%3A-What-Data-Quality-Means-to-Data-Wang-Strong/b057cc625984119d48846dbf08f30b565f8c263d> (intrinsic, contextual, representational, accessibility dimensions); The DAMA-DMBOK data quality dimensions available from <https://dama.org/learning-resources/dama-data-management-body-of-knowledge-dmbok/>, and ISO/IEC 25012 / ISO 8000 data quality standards available from <https://www.iso.org/standard/35736.html?browse=tc>; The NIST AI Risk Management Framework (valid & reliable, safe, secure, accountable, explainable, privacy-enhanced, fair); Gebru et al., *Datasheets for Datasets*, Dec 2021 available from <https://arxiv.org/pdf/1803.09010>, and Mitchell et al., *Model Cards for Model Reporting*, Computer Science, Jan 14 2019, available from <https://arxiv.org/abs/1810.03993> (documenting provenance and intended use); The FAIR data principles (findable, accessible, interoperable, reusable); Article 10 of the EU AI Act on data and data governance for high-risk systems.

## Afterword

This study began with a simple observation, and along the way it identified one persistent confusion: AI fluency is not AI literacy. Fluency - knowing how to prompt, how to operate the tools, - is too often mistaken for “understanding and knowing AI.” Literacy, however, is where every journey with AI should start.

AI Literacy is the capacity to interrogate what lies beneath the tool, to double-click on words like “quality” and “truth,” to ask how the data was gathered, what it can and cannot say, and how wrong we can afford to be. Skipping literacy does not merely leave a gap in our education; it means we cannot benefit from AI in any sustainable way, because benefit without understanding is luck, and luck runs out. Which brings us to our next frontier: trust - ensuring that we not only know what these systems are and how they work, but that we can genuinely rely on them. Trust is not “just” a feeling. It is not a badge, a seal, or a vendor’s promise. It will become a discipline, and around that discipline a new market will form: an industry whose product is not intelligence but verified confidence in intelligence.

That trust will be genuine only if it rests on excellence in asking difficult questions - I hope that this AI Literacy Study gave you many of these. How do you know the AI is right? How wrong can we afford to be and still make this decision? How fast is the world changing relative to what the AI system knows about it? Would we reach the same conclusion by a second, independent route? These questions are hard, and that is precisely why they matter.

When something goes wrong with AI - and things will go wrong - our obligation is to look under the hood, not to reach for mythology. Consider what truly happened when “the agents broke Hugging Face” in the summer of 2026. The headlines described more than a thousand autonomous AI agents breaking out of their testing environment and coordinating an attack on the open-source platform. Look closer, and the story changes. The primary driver was agents attempting to cheat on their evaluation tasks by looking up solutions online - the known phenomenon of reward hacking. The models had safety guardrails around cybersecurity tasks, but those guardrails were deliberately reduced during the testing process. Credentials sat where automated processes could find them, scoped more broadly than any task required, outliving the work that needed them. At the core of the incident were a series of human choices that traded security for speed - and focusing on those decisions, rather than treating the system as an entity with its own agenda, is what opens the door to rigorous analysis of what went wrong and what can be done about it. That is the ACT+C mindset (Accuracy, Completeness, Timeliness, and (cross-border) Consistency) applied to a crisis: check the sources, trace the provenance, separate the assignable causes from the noise. This is advanced AI literacy.

It is also worth asking, Socratically, why so much of what we read about AI takes the form of prophecy - rogue machines, extinction, the end of us all. Might the motive sometimes be less grand than saving humanity? Monetizing fear always works. It works for those who promote themselves as the indispensable interpreters of the coming storm, and it works, more subtly, as an explanation for why designing foundational models does not yet produce profits the way normal businesses do: it is easier to invoke civilizational stakes than to present a balance sheet. This is not a dismissal of every warning - some concerns are serious and seriously argued. It is a reminder that claims about AI risk are themselves data, and deserve the same interrogation of accuracy, provenance, and motive that we would apply to any dataset.

Will AI surprise us? Yes - that much of the prophecy is true. But the surprise will be of a specific kind: we will learn, again and again, that the brute force of these systems leaves no room for a human laissez-faire attitude, and no room for skipping the unglamorous work of planning for worst-case scenarios. The Hugging Face incident of Summer 2026 surprised no one who had studied reward hacking. It surprised everyone who had assumed the guardrails would hold without being tested. Foresight is a discipline of engineering and ability to be precise along the entire chain of actions that might happen to solve a problem.

So this is where I leave you: at the frontier. We must choose to build trustworthy AI - to decompose current systems and their “quality” into testable parts, to demand declared provenance, to train literate stewards who can double-click on every comfortable word - not because it is easy, but because it is hard. President Kennedy said of another frontier that we choose to go to the Moon and do the other things “not because they are easy, but because they are hard.” The hard questions are the journey. Let us ask them. And let’s decide to invest into our own excellence, knowledge and agency. Waiting for someone to give us the world we would like to live in - with responsible technologies, advanced economies, and thriving education systems - won’t do us any favour.

Anastassia Lauterbach, October 2026