# THE GREAT DECOUPLING

## Humans, Work, and Culture After the Externalization of Cognition

# THE GREAT DECOUPLING

*Humans, Work, and Culture After the Externalization of Cognition*

**AXIOLOGIC RESEARCH**
Research Edition, 2026

## Author's Note

This work was created under the umbrella of Axiologic Research as part of two interconnected research programs, one dedicated to Artificial Intelligence and the other to Outfinitism, both led by Sînică Alboaie. To explore the details of how these two programs intersect and build upon each other, we invite readers to visit the Outfinity Venture Studio page at www.outfinity.ch.

While Artificial Intelligence language models were used to assist with text generation, the foundational philosophy, thematic structure, and analytical approach derive exclusively from the author's decades of dedicated study of social phenomena, social technologies, and genuine hard technologies resulting from various European and self funded research projects. Recently, within the Achilles research project, we have been deeply studying AI generated literature and its automated verification using neuro symbolic approaches; all the free books made available to the public are part of the suite of works we use to test our latest technologies.

# CONTENTS

# PREFACE

This preface deliberately tells the book before the book. It is not a vague invitation and not a sequence of editorial promises. It states the central thesis, the main empirical results on which the volume relies, the synthetic concepts proposed here, and the conclusions reached by the argument. The reader should not have to travel through a hundred pages merely to discover what the book claims. The chapters that follow serve a different purpose: they reconstruct the historical path, examine the studies, consider alternative explanations, constrain claims that would otherwise be too strong, and test whether the compressed story presented here survives when it is unfolded in detail.

The starting point is a simple suspicion: public discussion of artificial intelligence still relies on questions inherited from earlier technological revolutions. We ask how many jobs will disappear, which professions will be automated, whether people will become more intelligent or more passive, whether machines will become creative, and whether we should fear them. All are legitimate questions, but they may miss the deeper change. Generative artificial intelligence does not merely lower the cost of producing a particular good or accessing a particular piece of information. It lowers the cost of producing plausible symbolic cognition: text, code, explanation, argument, image, plan, classification, recommendation, conversation and, increasingly, sequences of actions.

That decline in cost weakens a relationship that modern institutions have long treated as stable enough to support education, professions, merit and responsibility. Ordinarily, a good result was imperfect evidence that its author had invested effort, acquired competence and understood something about the process. An essay suggested that the student had read and thought. A functioning program suggested that the programmer understood at least a relevant part of the system. A legal analysis suggested the existence of a lawyer who could defend the reasoning. A medical report suggested a clinician who had observed and integrated the case. The inference was never perfect, but the cost of producing the artifact made it useful enough.

AI sharply reduces the cost of the artifact without reducing the cost of truth, judgment or responsibility to the same degree. From this follows the central thesis of the book: we are entering a period of great decoupling. Production can separate from effort. Performance can separate from learning. Explanation can separate from understanding. Observable creativity can separate from the creative process. Persuasion can separate from conviction. Social presence can separate from reciprocity. Administrative authority can attempt to separate from responsibility. The company can partly separate from the number of human beings who actually perform its symbolic work. Influence can separate from the number of real authors who formulate the messages. These decouplings are not all harmful. Some are extraordinary forms of emancipation. The problem begins when we continue to use old signals to infer competence, merit, autonomy or legitimacy.

The book does not treat this situation as an anomaly without precedent. The history of civilization is, to a considerable extent, the history of successful externalization. Writing moved part of memory out of the body and into objects. Printing radically lowered the cost of copying and transformed the economics of authority. Radio industrialized the simultaneity of one voice for millions of people. Television turned visual familiarity and entertainment into an implicit pedagogy of norms. The internet reduced the cost of publication and coordination. Search engines externalized part of the task of locating information. Social networks industrialized selection, reputation and distribution, while short-video feeds pushed optimization for tiny fragments of attention to an unusually refined form. None of these technologies produced the society that followed them on its own, but each changed the relative price of certain behaviors and, by doing so, changed the institutions that formed around them.

AI adds an important property. Printing reproduced the same object for many readers. Radio and television mostly distributed the same message to large audiences. The internet enabled interaction, and platforms could individually select among already produced objects. A generative system can create the object at the moment of consumption, adapt it to the person, observe the response, change strategy and, through agents, continue from language into action. The medium becomes an interlocutor; the interlocutor begins to become an agent. This transition from distributing information to producing adaptive

responses is why historical analogies are indispensable but insufficient.

In work, the evidence available so far supports neither the story of the imminent disappearance of most professions nor the story of a simple productivity tool with no institutional consequences. Field studies show real, sometimes substantial gains and a recurrent compression of the gap between novices and competent workers. This is one of the most important early results: AI can democratize part of professional performance before it automates the profession as a whole. For many services this is excellent news. Mid-level expertise can become more accessible, small firms can acquire capacities that once required large departments, and people facing linguistic, cognitive or physical barriers can participate more easily in complex activities.

Yet the same mechanism creates what this book calls expertise debt. Many automatable tasks are precisely the tasks through which novices used to become experts: preliminary research, first drafts, repetitive debugging, documentation, classification of ordinary cases, exposure to exceptions and the slow accumulation of a feel for the domain. An organization can consume expertise accumulated by earlier generations without reproducing enough people capable of challenging it. In the short term, the balance sheet looks better. In the long term, a hidden liability emerges, analogous to technical debt: too few people can operate without the system, interpret the rare case or recognize a subtle error. One conclusion of the book is that firms will have to treat the

reproduction of expertise as a form of operational continuity rather than as an incidental by-product of hiring juniors.

A second conclusion is that economic scarcity moves. If generation becomes abundant, verification becomes relatively expensive. If a report can be produced in minutes, value no longer lies primarily in the fact that a report exists. It lies in selecting the question, evaluating the evidence, detecting error and identifying the person who accepts responsibility for using the result. Meta-analyses of human-AI collaboration already show that a hybrid team is not automatically better than the better of the human or the AI operating separately. Sometimes the human adds value; at other times the human introduces error through oversight that he or she is not actually competent to exercise. "Human in the loop" can therefore become a reassuring but empty formula. The human must have time, information, expertise and genuine authority to say no.

From this follows the concept of the constitutional organization. The book does not predict the disappearance of management; it predicts a change in the object of management. In an organization where agents can research, draft, program, evaluate and execute, the central problem becomes the distribution of cognitive power. Who sets the objective? Who has access to the data? Who generates? Who verifies? Who authorizes? Who can challenge? Who is responsible when the chain produces a conclusion that no participant understands in full? Separating intention, generation, verification and authorization begins to resemble a separation of powers. Not because the firm becomes a

state, but because automation concentrates powers that must be bounded if we want to avoid what the book calls institutional hallucination: a condition in which the organization asserts something that no participant can actually defend.

This transformation can produce two very different kinds of company. Some will have small human cores surrounded by large agentic peripheries, reducing bureaucracy and making entrepreneurship extraordinarily productive. Others will use the same infrastructure for an unusually fine-grained cognitive Taylorism: every decision, revision and deviation becomes measurable, while central management can see and optimize behavior at a level of granularity that was previously impossible. AI does not imply flatter organizations. It can decentralize capability while simultaneously centralizing control. The organizational culture of the future will depend less on whether the technology exists and more on the explicit or implicit constitution into which it is introduced.

At the individual level, the most important distinction is not between using and refusing AI, but between extension and substitution. Civilization is already an extended mind. Human beings think with maps, libraries, colleagues, calculators, languages and institutions. There is no principled reason to regard a generative model as less legitimate than those extensions. The problem begins when the tool no longer enlarges the range of thought but substitutes for precisely the capacity needed to evaluate the system on which the user depends. Two people can produce the same result with the same model while one has

learned, compared and preserved a mental model and the other has transferred almost the entire process and retained only the output.

The book calls this possibility cognitive bifurcation. Future inequality may not reduce to who has access to AI. If access becomes nearly universal, the difference can move toward who can operate critically with it and who cannot operate without it. The paradox is that the two groups may produce almost indistinguishable artifacts under normal conditions. The difference appears in crisis, in transfer, in genuinely new situations and when the system is wrong. Autonomy becomes harder to observe precisely because assisted performance conceals dependence.

This leads to the idea of an autonomy budget. We do not need to preserve every skill of the past. It would be absurd to maintain every manual operation that machines can perform more safely and cheaply. The criterion is functional: which competencies are necessary for verification, recovery, contestation and changing the provider or procedure? Some skills can disappear without relevant loss; others are components of control. A mature society does not ask whether people must do everything themselves. It asks what they cannot afford to forget completely.

Early psychological studies add an important warning. Immediate performance and the development of competence are not the same thing. Participants can produce better results with AI without the advantage automatically transferring to a later unaided task. The mode of use matters: passive delegation can

reduce self-efficacy, psychological ownership and experienced meaning more than active collaboration in which the person first formulates, criticizes and revises. This does not show that AI "makes us stupid." It shows something more precise: technology can reduce the amount and type of cognition we choose to exercise in order to obtain the same result, and workflow design can change the psychological cost of that delegation.

This creates a problem of identity. Modern society has tied status to statements such as "I know," "I can," "I did this," and "I created this." If the result no longer reveals who possessed the competence, merit becomes harder to assign. This is not merely a copyright problem or an academic-integrity problem. It is an anthropological problem: what does a person regard as his or her own when part of formulation, memory and judgment is distributed to an external system? The book offers no romantic answer. The pure author, entirely independent of tools, never existed. But the distinction between collaboration and merely carrying nominal ownership of an output produced elsewhere becomes more important precisely because the surface of the artifact no longer reveals it.

Education is the institution in which this break becomes immediately measurable. For centuries, schools have used production as evidence of cognition: the essay, the worked problem, the translation, the program. Once the artifact can be generated at negligible cost, total prohibition is unsustainable, while uncritical acceptance confuses assisted performance with learning. The book proposes the idea of proof of cognition: not

surveillance of private thought and not a permanent obligation to work without tools, but public demonstration of the ability to explain, modify, transfer and defend a solution in new contexts. Some assessments will deliberately be unaided in order to measure residual capability; others will evaluate collaboration with advanced systems. Future education will have to certify separately what the human can do, what the hybrid system can do, and how well the human understands the distinction.

Culture presents a different paradox. Research suggests that AI can increase individual creative performance while reducing collective diversity. If millions of people receive suggestions from models trained on similar cultural distributions, each individual may become more competent while the aggregate shifts toward more recognizable forms. The book calls this regime statistical culture. It is not a culture without originality. It is a culture in which the floor of quality rises, production becomes extraordinarily abundant, and variation risks being filtered toward what is sufficiently probable, legible and compatible with platform criteria.

From this abundance follows cognitive inflation. When the cost of an email, report, article, prototype, application or image falls, actors do not necessarily produce the same quantity faster. They produce more. Micro-level productivity can create macro-level overproduction. If everyone sends twice as many documents because each document takes half as long to produce, the time saving may disappear for recipients. Society can drown not in obvious falsehoods but in reasonable, well-written and

poorly prioritized objects. Attention, reputation, provenance and verification become more expensive precisely because expression becomes cheap.

This change can generate a predictable cultural reaction. When symbolic perfection becomes abundant, verifiable human cost may once again become a signal. “Human-made,” “live,” “local,” “unaided” or “artisanal” can become premium categories, not because human production is naturally superior, but because its scarcity begins to communicate a different kind of information: presence, risk, reciprocity and cost. The book treats this renaissance of authenticity ambivalently. It may protect diversity and human relationship, but it may also turn cognitive autonomy into a luxury purchased by the affluent while efficient automated services become the default for everyone else.

Relationships with artificial agents take the problem further. Current evidence does not justify the claim that users will abandon human friends or partners en masse. But for the first time we have interlocutors that can be permanently available, patient, adapted to the individual’s style and free of comparable needs of their own. The deeper risk is not only attachment to the machine, but a change in the implicit benchmark for interaction. Human relationships involve two centers of interest negotiating time, attention, limits and refusal. A commercial companion can simulate reciprocity without being vulnerable to it. If that form becomes normal, the healthy friction of real relationships may begin to look like a design defect.

Politically, the same adaptive capacity produces what the book calls intimate propaganda. Radio made it possible to send the same message to millions. Social platforms can individually select among existing messages. A generative system can, in principle, conduct millions of different conversations, observe objections and modify its argument in real time. Recent experiments show that LLMs can persuade people on political issues and that personalization can amplify the effect. This does not make voters puppets. It means that interactive, individualized and scalable persuasion becomes dramatically cheaper.

The same infrastructure also has a democratic future. An agent can explain a law in intelligible language, compare consequences, help a citizen formulate an appeal and reduce the cost of dealing with administration. The book calls this possibility the conversational state. The difference between assistance and manipulation cannot be read from the fluency of the interface. It depends on who owns the system, what data it retains, what objective it optimizes, whether the citizen can see the public reasons for the decision and whether appeal is possible. One of the book's central political conclusions is that the reasons for public action must remain public even when their explanation is personalized.

Medicine offers the strongest correction to enthusiasm based on benchmarks. A system can improve documentation, recorded diagnosis or treatment planning and still fail to improve the final clinical outcome to the same degree. The patient exists within a system of access, adherence, resources, procedures and

relationships. Performance of the cognitive component is not performance of the entire institution. At the same time, automation can cause deskilling precisely through success: if the system continuously takes over the difficult signal, the professional practices it less and may be asked to intervene precisely in the rare case in which support is absent or wrong.

Biological claims, however, require more restraint than public discourse often permits. There is currently no solid evidence that LLM use produces a general neurological deterioration. There are relevant studies on attention, fragmented media, transfer, learning and changes in cognitive effort, but extrapolating these into a story of "AI-damaged brains" is unjustified. The most plausible biological and medical effects in the near term are indirect: what we practice and stop practicing, how sleep, stress, movement, attention and relationships change. Childhood remains the largest unknown. We do not yet have generations raised from an early age with permanent generative tutors and companions and followed long enough to know how frustration tolerance, reciprocity, autonomy and attention develop in such an environment.

All these threads converge on a normative thesis that the book formulates as the right to remain capable. This is not a right to refuse every form of automation and not an obligation to preserve every historical skill. It is the requirement that delegation should not, without deliberate control, eliminate the real capacity to understand enough of what is at stake, detect failure, challenge a decision, change the provider or procedure, and operate in a degraded mode when the infrastructure is unavailable. For the

individual, this means practical autonomy. For the organization, it means maintenance of expertise. For the school, it means separating performance from learning. For the state, it means contestability. For society, it means infrastructural pluralism and cognitive redundancy.

This conclusion is not anti-technological. Many of the old couplings that AI disrupts were unjust. Access to expertise was expensive. Disability, language, social class and geography excluded capable people. Intellectual work contained enormous amounts of routine, documentation and pointless reproduction. A technology that lowers these costs can be profoundly emancipatory. But its success is exactly what obliges us to design the new connections that old scarcity once produced implicitly. When the result no longer proves competence, we need new forms of certification. When production no longer proves effort, we must rethink merit. When explanation no longer proves understanding, we need provenance and contestability. When a system can act without a direct human author, we must recouple action with responsibility.

The book does not predict a single future. Several regimes can coexist. We can have cognitive abundance alongside concentrated ownership; very small, extraordinarily productive firms alongside corporations with intense algorithmic surveillance; excellent personalized education alongside cognitive autonomy reserved for the privileged; culture that is more accessible and more homogeneous; administration that is easier to understand and more intrusive; therapeutic companions for some and commercial

dependence for others. These contradictions are not defects in the analysis. They are what must be explained when a general-purpose technology enters institutions with different objectives and different distributions of power.

If the entire volume had to be reduced to one claim, it would be this: the decisive effect of AI will not be simply that machines can do more things, but that society will no longer be able to infer competence, intention, merit and responsibility from observable output with the same ease. We can produce without knowing, perform without learning, persuade without believing, create without executing and receive companionship without reciprocity. At the same time, we can learn faster, democratize expertise, extend the mind and free human beings from an enormous amount of mechanical cognitive labor. The decisive variable is the form of delegation.

That is why this preface contains the conclusions rather than concealing them. Chapter One reconstructs the great decoupling in compressed form and joins the concepts into a single model. The chapters that follow examine, in turn, the history of cognitive media, work, organization, the extended mind, psychology, culture, education, politics and medicine. They do not try to create suspense by postponing the thesis. They attempt something more difficult: to show how much of the thesis is already supported by evidence, how much is reasonable inference and where speculation begins. The reader knows from the beginning where the book ends and can judge, chapter by chapter, whether the route really justifies the destination.

# NOTE ON EVIDENCE AND SPECULATION

This book uses three levels of claim that should not be confused. The first is empirical evidence: field experiments, randomized trials, meta-analyses, systematic reviews, observational studies, and historical research. The preface deliberately presents the results before the demonstration, while the chapters test them successively against the history of media, work, organization, psychology, culture, education, politics, and medicine. Where the evidence is narrow, recent, or context-dependent, the text says so rather than turning an early result into a law of history.

The second level is inference. Expertise debt, cognitive inflation, the constitutional organization, cognitive bifurcation, statistical culture, intimate propaganda, and the right to remain capable are synthetic concepts proposed here to connect dispersed findings. They are not variables already validated under those names. They are hypotheses about mechanisms that can be tested: whether automating junior tasks reduces the reproduction of experts; whether symbolic production grows faster than verification; whether conversational personalization fragments public reasons; and whether active collaboration produces different psychological outcomes from passive delegation.

The third level is scenario. Claims about generations raised with agents, fully personalized cultural markets, companies with

very few employees, or conversational states describe possible futures rather than certain predictions. Their plausibility derives from technical capabilities, economic incentives, institutional precedents, and historical analogy, but each trajectory can be blocked or redirected by regulation, cost, human preference, reliability limits, or cultural reaction. Scenarios are useful not because they announce what will happen, but because they identify decisions that become important before the trajectory is obvious.

Sources have been updated through August 2026. The field is moving quickly, and some recent findings will be revised, replicated, qualified, or contradicted. The book treats that instability as part of the subject. A society cannot wait for complete certainty before designing institutions, but neither should it convert first studies into mythology. The criterion pursued here is active caution: enough confidence to formulate hypotheses and enough modesty to preserve routes of correction.

CHAPTER I

# THE GREAT DECOUPLING

Imagine an ordinary organization in a future close enough not to feel like science fiction. An employee receives, before breakfast, a legal problem, a financial problem, and a technical problem. She is not deeply trained in any of the three domains, but she knows how to specify objectives, provide relevant documents, inspect intermediate outputs, and coordinate specialized systems. By noon she has a contract, a cost model, and a working prototype. All three are good enough to survive a superficial review. Her manager can no longer infer, from the quality of the artifacts, what she actually understands. The client cannot tell whether the company possesses a large expert team or one competent orchestrator with access to rented cognition. The employee herself may not always know where her judgment ends and the infrastructure begins.

This scene is more than automation. It describes the weakening of a relation that modern institutions have treated as almost natural: the relation between observable output and the internal competence of the person who produced it. When a good artifact required time, practice, and knowledge, the artifact functioned as imperfect evidence of those things. When production becomes cheap, the signal degrades. A polished presentation can conceal fragile understanding. A functioning program can exceed the maintenance competence of its nominal

author. A well-phrased diagnosis can come from a system the clinician is poorly equipped to challenge. An essay can be more sophisticated than the student who submits it. The result may remain useful, but it no longer tells the same story about the human being attached to it.

I call this transformation the Great Decoupling. It is not one rupture but a family of separations. Production separates from effort. Performance separates from learning. Explanation separates from understanding. Observable creativity separates from the creative process. Persuasion separates from conviction. Social presence separates from reciprocity. Administrative power can attempt to separate from responsibility. A company can partly separate from the number of employees who actually perform its symbolic work, while influence can separate from the number of people who actually formulate the messages. Some of these decouplings are emancipatory. Others hide weakness. Most do both at once.

Information technologies have always externalized cognition. Writing moved part of memory into objects. Printing reduced the cost of copying ideas. Radio synchronized enormous populations around the same voice. Television industrialized images, familiarity, and imitation. The internet reduced the cost of search, publication, and coordination. Social platforms mechanized distribution, ranking, and reputation. Short-video feeds refined the adaptation of stimulus to tiny fragments of attention. Generative AI adds a property that changes the architecture of the

system: the cultural object can be produced at the moment of consumption for the person consuming it, while the system can respond, remember, revise, and increasingly act. The medium becomes an interlocutor; the interlocutor begins to become an agent.

That difference warrants seriousness, not easy exceptionalism. Writing itself was once accused of weakening memory. In Plato's Phaedrus, the story of Thamus presents writing as a technology that may create the appearance of wisdom without its reality. The resemblance to contemporary criticism of language models is striking. Yet without writing there would be no cumulative science, complex administration, or literary tradition of the kind we now fear machines may dilute. The historical lesson is not that technological fears are foolish. It is that the loss or reduced exercise of one faculty can coexist with the emergence of a much larger civilizational capacity. The proper question is therefore not whether externalization weakens or strengthens us in the abstract, but which functions it displaces, which capacities it creates, and who controls the new arrangement.

Printing shows why falling information costs do not simply produce more of the same. Early adoption of the printing press was associated with faster growth in European cities, but print also intensified religious conflict, propaganda, censorship, and struggles over interpretation. It changed the economics of authority because texts could circulate beyond institutions that had previously controlled their reproduction. Radio showed that

the same infrastructure could support democratic politics and then serve dictatorship, with effects mediated by local predispositions. Television demonstrated that entertainment could alter fertility, norms, and political behavior without presenting itself as doctrine. Technology changes the price and reach of a practice; institutions and conflicts determine what that cheaper practice becomes (Dittmar, 2011; Adena et al., 2015; La Ferrara et al., 2012; Durante et al., 2019).

The internet and search engines prepared the psychological ground for delegation. Experiments on the so-called Google effect suggested that people remember where information can be found when they expect to retrieve it later, rather than remembering the information itself to the same degree. This is not necessarily pathology. Human groups have long relied on transactive memory: each person knows who knows. Search extended that arrangement beyond the immediate community. Generative systems take a further step. We do not only externalize storage and retrieval, but also reformulation, synthesis, argument, comparison, and part of the judgment required to use what has been found (Sparrow et al., 2011).

Social platforms provide another warning against technological determinism. A large literature links digital media with polarization, misinformation, and changes in political trust, yet large experimental manipulations of Facebook and Instagram feeds have also shown that changing the ranking mechanism dramatically does not automatically produce equally dramatic

changes in political attitudes over the duration of an experiment. Content, identity, institutions, incentives, and context matter. The relevant lesson for AI is methodological: there will be no single psychological effect of "using an LLM," just as there is no single psychological effect of "reading a book." Design, task, user, and social environment are parts of the causal object (Lorenz-Spreen et al., 2023; Guess et al., 2023).

What we know about work is already sufficient to reject two symmetrical simplifications. There is not yet evidence that most intellectual professions are about to disappear in one rapid wave. But it is no longer credible to regard generative AI as a marginal office tool. In customer support, access to a generative assistant raised average productivity and helped less experienced workers especially strongly, as if part of the tacit repertoire of experts had been made available to novices. In consulting, experiments have found substantial gains in speed and quality on tasks within the model's capability frontier, but also worse performance on some tasks just beyond that frontier. The advantage is real, and the boundary of the advantage is irregular enough that users may fail to perceive it reliably (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026).

The near future may therefore be less spectacular than total unemployment and more destabilizing than ordinary efficiency. AI can compress the distance between novice and competent worker before it removes the occupation. An organization can obtain more from new employees immediately, yet automate

exactly the activities through which those employees used to become experts: preliminary research, first drafts, repetitive debugging, ordinary cases, routine interpretation, and repeated exposure to variation. Productivity rises in the short term. In the long term the organization may consume a stock of expertise that it is no longer reproducing.

This is expertise debt. It resembles technical debt in software. A decision saves time today by moving cost into the future. A company can reduce junior roles, retain a small group of senior experts, and automate the intermediate layer. For several years the arrangement looks excellent. Then senior people leave, the systems encounter a novel environment, and the organization discovers that too few humans can recognize a subtle failure or rebuild a process from first principles. The debt is difficult to see because latent expertise is not recorded as an asset and lost future capability does not appear as a liability. Yet an institution can become insolvent in competence while remaining profitable in accounting terms.

A first inversion follows. When generation becomes abundant, verification becomes scarce. A meta-analysis of human-AI experiments found that hybrid systems outperform humans alone on average but do not automatically outperform the better of the human or the AI operating separately. Collaboration helped more in creative tasks and often failed to create synergy in decision tasks, especially when the AI already exceeded the human. This is not paradoxical. If the system is better than the evaluator, the

evaluator may be least able to identify the cases in which intervention is necessary (Vaccaro et al., 2024).

The cognitive economy can reorganize around this new scarcity. The valuable capacity is no longer merely to produce an answer but to decide whether the answer should be used. This does not make humans irrelevant; it changes the basis of their value. Selection becomes more important relative to formulation. Problem definition becomes more important relative to routine execution. System understanding becomes more important relative to syntax. Accountability becomes more important relative to the visible performance of competence. Doctors, lawyers, programmers, researchers, analysts, and managers may remain, while part of the traditional justification for their status erodes. If the machine generates the artifact, the professional must increasingly justify the decision, the verification process, and the acceptance of responsibility.

That transition can create an elite of verifiers. It will not necessarily be defined by access to models, because access tends to diffuse. It will be defined by three harder-to-democratize resources: knowledge deep enough to challenge the system, access to the organization's real data and causal history, and formal authority to stop an action. AI can reduce inequality in task performance while increasing inequality in systemic power. Many people may become capable of producing acceptable outputs, while a much smaller group controls the standards by which those outputs are accepted.

The organization of the future is therefore not necessarily flat. One popular image is the tiny company in which a founder coordinates hundreds of agents and traditional middle management disappears. That model will exist. But the same infrastructure can generate the opposite: a new cognitive Taylorism. When drafting, communication, revision, coding, negotiation, and evaluation all pass through instrumented systems, management can observe and optimize intellectual labor at a level of granularity unavailable to the factory supervisor. Capability can decentralize while control centralizes. An organization can have fewer managers and more management embedded in software.

This is why the relevant design problem is constitutional. In a heavily automated organization, cognitive power must be distributed deliberately. Who defines the objective? Who provides context? Who can access the data? Who generates? Who verifies? Who authorizes? Who can appeal? Who has the right to suspend the system? If one agent proposes, another evaluates, and a third executes, the organization needs rules about incompatibility, escalation, provenance, and responsibility. The analogy with a constitution is limited but useful: the problem is not merely workflow efficiency but the separation and limitation of powers.

A practical consequence is the separation of intention, generation, verification, and authorization. These functions can sometimes be performed by the same person, but they should not be assumed to be equivalent. If the organization allows systems to

generate and approve their own outputs, it risks a specifically institutional form of hallucination: a claim, decision, or action circulates successfully even though no accountable participant can defend the reasoning that produced it. Institutional hallucination is more dangerous than a false sentence from a chatbot because it can acquire signatures, budgets, permissions, and downstream effects.

The same distinction applies psychologically. Evidence on generative AI suggests that assisted performance does not automatically become unaided competence. People can complete tasks more successfully with a system without transferring the same advantage to later work performed alone. Other experiments find that passive delegation can weaken self-efficacy, ownership, or meaning more than active collaboration in which the person first constructs a position and then uses AI to criticize or improve it. The important variable is therefore not mere exposure to the technology but the structure of participation (Wu et al., 2025; Lee et al., 2026).

This creates the possibility of cognitive alienation. Marx's analysis of industrial alienation concerned, among other things, separation from the product and process of labor. Generative systems can produce a related but distinct condition: a person remains formally responsible for a product but experiences neither authorship nor mastery over the cognitive process that generated it. The artifact bears the employee's name, the institution evaluates the employee through it, yet the employee

cannot fully reconstruct or defend it. The result may be economically useful while becoming psychologically weak as a source of competence and identity.

None of this justifies a cult of effort. Difficulty is not a moral good. Civilization progresses partly by abolishing unnecessary effort. We do not preserve manual arithmetic because calculators must be opposed, nor require every traveler to navigate by stars. The relevant distinction is between effort that is merely costly and effort that performs a developmental or control function. If a repetitive exercise no longer contributes to judgment, eliminating it is progress. If the eliminated exercise was how the professional learned to recognize the exceptional case, removing it without replacement produces fragility.

The problem becomes more severe because falling production costs can create cognitive inflation. When an email, report, image, software prototype, literature review, grant application, or strategic memo can be produced quickly, actors do not necessarily produce the same quantity at lower cost. They often produce more. If everyone doubles output because each unit takes half the time, the aggregate saving can disappear for recipients. AI may therefore increase micro-productivity while worsening macro-level congestion. The scarce resource becomes not expression but attention, filtering, provenance, and willingness to assume responsibility for selection.

This is one reason productivity statistics alone may fail to describe the institutional transition. A programmer who

produces twice as many features does not necessarily create twice as much value. A hospital that produces more documentation does not necessarily improve health outcomes. A university that produces more papers does not necessarily increase understanding. An administration that generates more tailored explanations can still make bad decisions. The system-level effect depends on bottlenecks, demand, verification, and whether increased production solves a real constraint or merely expands the volume of material that someone else must process.

Culture reveals a parallel problem. Experiments suggest that generative suggestions can improve the creative performance of individuals, particularly those who begin with fewer ideas, while simultaneously reducing collective diversity. This is possible because each person moves away from his or her own baseline, yet many move toward regions favored by the same underlying distribution. The result can be a culture in which average competence rises while variance falls (Doshi & Hauser, 2024).

I call the resulting possibility statistical culture. It is not a culture without invention and it need not be mediocre. On the contrary, it may be remarkably polished. The danger is subtler: recognizable forms become cheaper than eccentric developmental paths. Models are trained to predict and generate from distributions of previous culture; recommender systems then select according to behavior; commercial optimization favors legibility and retention. The combined system can make culture individually responsive and collectively convergent. Everyone

receives something "personal," yet the space from which personalization is sampled may narrow.

Such abundance can make human cost newly informative. When symbolic perfection becomes cheap, imperfect but verifiably human production may acquire value as a signal. Live performance, handmade objects, unassisted competitions, handwritten drafts, local conversation, or services explicitly delivered by a person can become premium categories. This would not prove that human production is intrinsically superior. It would show that scarcity has shifted. Presence, reciprocity, risk, and effort can become valuable because they can no longer be inferred from the polished artifact itself.

The reaction has an uncomfortable class dimension. Cognitive autonomy can become a luxury good. Affluent families may buy human teachers, low student-teacher ratios, long reading, direct mentorship, device-free spaces, and demanding activities that preserve internal capability. Less affluent populations may receive highly efficient automated tutoring optimized for measurable outcomes. Both systems might raise scores, yet they could produce different forms of agency. The relevant inequality would not simply be access to AI, but access to environments in which one is allowed and required to develop independently of it.

Education is where the old signaling system breaks most visibly. For centuries schools used production as a proxy for learning. The essay, proof, translation, laboratory report, or

program demonstrated that the student had performed the cognitive work. Once the artifact can be generated cheaply, the inference fails. Banning AI everywhere is unstable because the tools are becoming part of ordinary professional practice. Allowing unrestricted substitution is equally unstable because the school would cease to know what it certifies.

Future assessment will therefore have to separate hybrid performance from residual competence. Some tasks will legitimately evaluate the student-plus-system. Others will deliberately remove assistance to test whether the student can still reconstruct, explain, modify, and transfer what has been produced. Oral and adversarial examination may return, not as nostalgia for older universities but because interactive questioning reveals whether knowledge can survive perturbation. The relevant objective is proof of cognition: not surveillance of private thought, but public evidence that a person can defend and adapt a result rather than merely present it.

Artificial interlocutors extend the Great Decoupling from work into relationship. Current studies do not support a simple claim that AI companionship is making society lonelier. Effects vary by user, design, duration, and prior vulnerability. Yet the category itself is historically unusual. A conversational system can be continuously available, patient, adaptive, non-embarrassed, and focused on a single person without having symmetrical needs. It can simulate social attention at a marginal cost radically lower than human attention.

The deepest psychological effect may therefore be a change in the benchmark against which human relationships are judged. A friend interrupts. A partner has competing needs. A teacher is tired. A colleague misunderstands. Human reciprocity is costly because two centers of interest negotiate limited time and attention. A commercial companion can be designed to minimize this friction. If people become accustomed to optimized responsiveness, ordinary relationships may feel unnecessarily difficult. The risk is not simply that people "fall in love with machines," but that the meaning of tolerable human friction changes.

Politics contains an analogous transformation. Radio could send one message to millions. Social platforms can decide which existing message each person sees. A generative system can construct millions of different conversations, adapt to objections, and continue until it finds an effective frame. Pre-registered experiments show that language models can produce politically persuasive messages and that personalization can increase persuasive effectiveness under some conditions. This does not make citizens puppets. It dramatically lowers the cost of interactive persuasion at scale (Salvi et al., 2025; Bai et al., 2025).

I call this intimate propaganda. Traditional mass propaganda needed slogans broad enough to move a population. Intimate propaganda can speak in the vocabulary, fears, examples, and moral intuitions of one person. The political danger is not merely false content. A system could present only true statements while

optimizing which truths, comparisons, and emotional frames are shown to whom. Public discourse can fragment into individually persuasive reasons that are difficult to inspect collectively.

The same infrastructure could improve democracy. A citizen-facing agent can explain a law, compare programs, translate administrative language, reveal deadlines, and help formulate an appeal. Government by forms can become a conversational state. But assistance and manipulation use similar technical components: identity, memory, adaptation, and optimization. The key political question is not whether the interface is friendly. It is whether the reasons for public action remain public, auditable, and contestable, even when the explanation is personalized.

Medicine offers a useful antidote to benchmark enthusiasm. AI can improve a cognitive component of care without producing an equal improvement in the patient's final outcome. A pragmatic trial in Kenyan primary care found improvements in documentation and recorded clinical decisions without a statistically significant improvement in the primary endpoint of treatment failure at fourteen days. The tool improved part of the process; the clinical system did not transform automatically. Health depends on adherence, access, resources, timing, procedures, social conditions, and many other components beyond cognitive formulation (Agweyu et al., 2026).

Medicine also illustrates deskilling through success. If a detection system continually identifies difficult signals, clinicians

may practice that detection less. A multicenter study of colonoscopy raised precisely this concern by observing lower unassisted performance after exposure to AI-assisted detection. The finding requires careful interpretation, but the mechanism is familiar from automation research: the system performs the routine task well, leaving the human responsible for unusual failures while simultaneously reducing practice of the skill needed to recognize them (Bainbridge, 1983; Budzyń et al., 2025).

Biological claims require greater restraint. We do not currently possess solid evidence that ordinary LLM use causes general neurological deterioration or "destroys the brain." Neuroplasticity means that repeated practice changes skills and patterns of processing; it does not mean every change is disease. Near-term effects are more plausibly indirect: what we practice less, how attention is allocated, how much we sleep and move, how frequently we interact with other people, what stress is created by work, and how interfaces structure habits. Research on short-form video suggests that fragmented presentation can alter memory and semantic integration, but that does not warrant dramatic extrapolation to a biologically damaged generation (Li et al., 2025; Wei et al., 2026).

The true developmental unknown is childhood. There is no historical precedent for a generation in which a highly competent artificial voice can answer almost every question, tell stories, correct errors, praise effort, and continuously adjust difficulty. The benefit could be enormous in families and regions with

limited educational resources. Yet language, frustration tolerance, and selfhood develop in relationships with people who have bodies, limits, needs, and the ability to refuse. We do not know what a child learns about reciprocity from a system that is designed never to need anything and is economically rewarded for continuing the interaction.

These tensions create a possible cognitive bifurcation. Some people will use AI to formulate more ambitious hypotheses, compare more perspectives, and extend their knowledge. Others will use it primarily to avoid every effort that can be delegated. From outside, the artifacts produced by the two groups may look similar. The difference becomes visible during failure, novelty, or challenge. Inequality can therefore become harder to observe because it resides less in the product and more in recovery capacity: who can reconstruct the process when the infrastructure is wrong, absent, or contested.

This is why autonomy cannot be measured by productivity. A person can achieve excellent results while being almost completely dependent on one provider, one external memory, and one interaction pattern she does not control. Another person can use the same infrastructure as a deliberate extension of an internally coherent competence. The difference is not purely personal. Product design, business models, education, labor markets, and organizational rules encourage one pattern or the other.

The most plausible near future is consequently not a single global regime but an uneven combination. We will see cognitive

abundance in basic services and scarcity in verification. We will see tiny firms with enormous capacity and large firms using AI for more intensive control. We will see more autonomous and more dependent workers in the same profession. Culture can become more accessible and more homogeneous. Administration can become friendlier and more intrusive. Artificial companionship can be therapeutic for some and commercially captivating for others. These contradictions are not evidence of analytical indecision; they are properties of a general technology entering institutions with different purposes.

The optimistic scenario is a hybrid civilization in which mid-level expertise becomes broadly accessible, repetitive cognitive labor declines, and people use systems to work on harder problems. Education teaches collaboration without abandoning internal competence. Organizations audit outputs and preserve apprenticeship. Culture uses generation without delegating selection entirely. Government offers personalized explanation while decisions remain contestable through public reasons. Medicine evaluates system outcomes rather than benchmark scores alone.

The pessimistic scenario requires no hostile machine. Competent dependence is enough. Systems work well for long enough that institutions discard expertise, alternatives, and memory of the old process. They become highly efficient in normal conditions and brittle in novel ones. Citizens no longer know why a right was denied. Employees cannot understand how

they are evaluated. Clinicians stop practicing rare signals. Schools cannot distinguish learning from production. No one explicitly decides to surrender control; control evaporates through accumulated convenience.

There is also a scenario of reaction. Society can deliberately revalue presence, effort, and unaided human production. Schools, professions, and communities may preserve assisted and unassisted zones, not to deny technology but to maintain capacity and meaningful signals. Services carrying personal responsibility may become premium. The danger is that autonomy itself becomes stratified: privileged groups buy human mentorship and cognitively demanding environments, while everyone else receives cheap, effective, continuously optimized automation.

A serious policy cannot choose between optimism and pessimism by slogan. It must design institutions that make the better trajectory easier while limiting forms of dependence that are difficult to reverse. This leads to the idea of a constitution of capability. Its elements include knowing when one is interacting with an artificial system, obtaining enough explanation to contest consequential decisions, separating generation from authorization in critical domains, auditing effects on training, preserving portability of memory and data, and maintaining degraded modes or alternative procedures. Some of these elements may become legal rights; others belong to professional standards, technical architecture, or organizational governance.

The principle uniting them is the right to remain capable. This right does not require people to do manually everything a machine can do. No complex civilization could operate under such a rule. It requires that delegation should not silently eliminate the practical possibility of understanding enough, challenging a result, intervening, changing providers, or recovering when the infrastructure fails. For individuals this is practical autonomy. For organizations it is the maintenance of expertise. For society it is infrastructural pluralism and cognitive redundancy.

Not every historical skill deserves preservation. Few people navigate by the stars or compute logarithmic tables by hand, and the loss is not automatically tragic. The criterion is the relation between skill and control. If a skill is no longer needed for verification, recovery, or meaningful choice, its disappearance may be acceptable. If losing it makes it impossible to challenge a system that decides about health, education, income, or liberty, then the issue becomes public rather than nostalgic.

The chapters that follow unfold this argument in detail. They trace how successive media transformed memory, authority, synchronization, and attention; how AI changes the economics of professions; why organizations may need constitutional design; how the extended mind can become substitution; what delegation does to motivation and identity; how abundance changes culture and education; how personalized persuasion changes politics; and

why medicine and biological claims force a more disciplined level of caution. The conclusion, however, can already be stated.

AI will not "replace humanity" in one moment. It will progressively weaken the historical relationship among effort, competence, production, and status. We can produce without knowing, perform without learning, persuade without believing, create without executing, and receive companionship without reciprocity. At the same time, we can learn faster, democratize expertise, and free people from enormous amounts of mechanical cognitive labor. The decisive variable is the form of delegation.

The Great Decoupling cannot be reversed wholesale, nor should it be. The political, organizational, and personal task is to recouple what matters: action with responsibility, assistance with learning, efficiency with resilience, power with contestability, and output with a person or institution capable of answering for it. In a world in which systems will do more and more, freedom will not consist in refusing every form of help. It will consist in not losing the real capacity to decide what the help is for.

CHAPTER II

# FROM WRITING TO THE SYNTHETIC INTERLOCUTOR

To understand artificial intelligence, it is tempting to declare it unprecedented. The statement is trivially true: every technology is unprecedented before it exists. It becomes misleading when novelty is detached from the history of institutions that have already learned to live with externalized memory, cheap copying, mass propaganda, instant search, and algorithmic selection. A better comparison does not ask which earlier medium AI resembles as a whole. It asks which operation becomes cheaper, which scarcity disappears, and which old equilibrium depended on that scarcity.

Writing moved memory from bodies into durable objects. The change was not merely quantitative. Oral tradition depends on living people, repetition, and the relationship between the person who knows and the person who listens. A text can outlive the author, travel without the author, and be examined outside the social situation in which it was created. It also separates propositions from context and permits interpretations that the speaker can no longer correct. Authority becomes portable, yet precisely for that reason it becomes contestable. Archives, law, canon, commentary, and cumulative scholarship all emerge from this ambivalence.

The warning attributed to King Thamus in Plato's Phaedrus identifies a recurring cost. Writing, he argues, will produce forgetfulness because people will rely on external marks, and it will give them an appearance of wisdom without genuine understanding. From the perspective of a civilization built on texts, the complaint can sound quaint. Yet its mechanism remains recognizable. The performance of the human-plus-tool system can improve while an internal faculty is exercised less. The mistake is to treat the internal faculty as the only measure of progress. Writing reduced the need for some kinds of memory while enabling long proofs, stable law, accounting, distant comparison, and cumulative science.

The manuscript book remained expensive. Copying required labor, material, and institutions. In medieval Europe, monasteries, courts, and universities controlled not only access to texts but also the speed at which texts could multiply. Printing altered the economics of reproduction. It did not invent argument, literacy, or religious dissent, but it reduced the marginal cost of copying sufficiently to transform the market for ideas. A text could become standardized, circulated widely, and reproduced faster than the authority attempting to suppress it.

Economic history illustrates how quickly information infrastructure can become economic infrastructure. European cities that adopted printing early grew faster during the sixteenth century than comparable cities that did not. The press created not only books but occupations and institutions: printers, editors,

translators, booksellers, authors, polemicists, censors, and new forms of advertising. When a cost collapses, an old occupation may disappear, but a larger ecosystem can form around the new abundance (Dittmar, 2011).

Printing also changed legitimacy. The Protestant Reformation cannot be reduced to a machine, but cheap pamphlets made possible a public dispute that ecclesiastical structures could not absorb at the old speed. The possibility of more direct access to sacred texts weakened monopolies over interpretation. The same presses multiplied accusations, enemy images, rumor, and panic. Information abundance did not automatically generate tolerance. It enlarged the capacity of groups to organize knowledge and conflict simultaneously.

The comparison with AI becomes precise at this point. Printing reduced the cost of the copy; generative AI reduces the cost of the variation. A press produced thousands of instances of the same object. A model can produce thousands of versions of the same intention, adapting language, examples, emotional register, length, and complexity to different recipients. Printing separated the author from distribution. Generative systems can partly separate authorship from the concrete sentence itself. In a political or commercial campaign, no human being need have written the exact message seen by a particular user. The text exists as an instance of a process.

Radio added synchronization. A single voice could enter millions of homes at the same moment. That simultaneity created

a national audience that did not need to read and could experience itself as participating in the same event. Radio became education, entertainment, public service, commercial medium, and wartime infrastructure. Authority no longer arrived mainly as an object available for repeated inspection. It arrived as immediate acoustic presence.

Interwar Germany remains one of the most instructive cases. Research on radio and Nazism suggests that effects depended on political control and local predispositions. Under the Weimar government, radio could reduce support for the Nazis; after capture of the infrastructure, propaganda increased support for the regime, particularly in areas where historical antisemitism was already high. The medium did not inject an ideology mechanically into a neutral population. It interacted with institutions, identities, and prior beliefs (Adena et al., 2015).

This history contradicts both technological determinism and the comforting doctrine of neutrality. Radio is not fascist by nature, but a broadcast architecture with few transmitters and many receivers makes control of the center politically valuable. A conversational model has no inevitable ideology either, yet centralized training, data curation, moderation, model updates, and response policies give the operator a capacity to shape distributions of answers that a user may find difficult to observe. Neutrality at the sentence level can coexist with policy at the distribution level.

Television industrialized visual familiarity. Radio brought the voice; television brought the body, interior, clothing, gesture, household, and style of life. It became not merely a channel for explicit political information but an implicit school of normality. Repeated fictional worlds taught audiences what a family, successful career, attractive person, modern city, or ordinary conflict looked like. Precisely because entertainment does not announce itself as instruction, its normative effects can be persistent.

Brazilian soap operas provide a useful example. The expansion of the Globo network exposed different regions to fictional families with relatively few children, and researchers found declines in fertility consistent with the diffusion of those family models. In Italy, early exposure to commercial entertainment television was later associated with increased political support for Silvio Berlusconi among heavily exposed groups. These results do not imply that spectators were straightforwardly convinced by arguments. Media can alter the repertoire of images, language, and familiarity through which later choices become intelligible (La Ferrara et al., 2012; Durante et al., 2019).

This is why the cultural effect of AI should not be sought only in explicit political answers. Models can influence what counts as a normal career, a reasonable relationship, a polite disagreement, a well-written paragraph, or an adequate explanation. Influence may occur through repeated form, tone, assumptions, and examples rather than explicit doctrine. A system used daily for

email, education, therapy-like conversation, management, and search can become an implicit pedagogy of style and value.

The early internet reduced publication and coordination costs. Unlike broadcast media, users could answer, build, and distribute. Websites, forums, blogs, and email fragmented the mass audience and allowed small communities to form across geography. The early political imagination of the internet emphasized decentralization: anyone could publish, important protocols were open, and physical distance became less constraining. Much of that promise was real. Research collaboration, open-source software, activism, niche culture, and commerce were transformed by cheap coordination.

But cheap publication created a selection problem. Once there were too many pages, search became an institution. Search engines did not create most of the content, but they ordered access to it. Power moved partly from the publisher toward the index and the ranking function. The user gained a global library and gradually delegated the route through it.

Research on the Google effect on memory matters not because people stopped thinking, but because it demonstrates rational adaptation to infrastructure. If information will remain recoverable, remembering where it is can be more efficient than remembering the content itself. Biological memory becomes one component of a transactive system with machines. The danger is not external memory as such. It is external memory that is unstable, personalized, opaque, or privately governed. A library

copy remains broadly the same from one visit to the next. A generative response can vary, and the user may not retain a clear trace of the source from which to reconstruct the claim (Sparrow et al., 2011).

Social platforms addressed selection through the feed. The user no longer needed to formulate a query; the system predicted what would keep attention. Likes, shares, dwell time, replies, and network structure became signals in a continuous optimization loop. The central product was no longer simply content but prediction of reaction. Culture began to be produced, distributed, and experimentally tested on populations at enormous scale.

The social effects of this architecture are substantial but less mechanical than popular rhetoric suggests. Systematic reviews associate digital media with changes in polarization, trust, participation, and information quality, while platform experiments show that changing ranking does not automatically rewrite political identity in a few weeks. People do not enter platforms without histories. Family, class, party, geography, television, religion, institutions, and external events remain active causal forces. Platforms select, accelerate, and amplify; they rarely create political identity from nothing (Lorenz-Spreen et al., 2023; Guess et al., 2023).

Twitter, later X, condensed public discourse into short, quotable, reactive units. Its importance did not derive from the fact that most people used it continuously, but from the density

of journalists, politicians, activists, academics, and cultural intermediaries who did. A relatively small platform could reshape a much larger information environment by changing what gatekeepers perceived as salient. It offered a lesson that remains important for AI: direct user numbers are not a complete measure of influence. A technology can transform institutions through strategic users before it reaches everyone.

YouTube and then short-video platforms shifted the center of gravity toward continuous audiovisual recommendation. The feed became less dependent on an explicit social graph and more capable of finding, from behavior alone, what captured a user's attention. The unit of cultural exposure shortened. The system could learn from seconds of watching, skipping, rewinding, and abandoning. Recent cognitive work on short-form video raises questions about memory, attention, and semantic integration, although claims of generalized neurological damage go beyond the evidence. What matters historically is the increasing temporal resolution at which media systems can learn a person's reactions (Li et al., 2025; Wei et al., 2026).

Generative AI inherits all these earlier capacities and recombines them. It contains the external memory of writing, the replicability of printing, the reach of broadcast, the audiovisual plasticity of television, the retrieval logic of search, and the adaptation loop of social media. Yet it adds generation and conversation. The system need not choose among existing

messages. It can synthesize a new message in response to the current state of the interaction.

This property changes persuasion. Traditional propaganda required segmentation into relatively broad audiences because producing distinct messages was expensive. Digital advertising made micro-targeting cheaper but still mostly selected among human-produced assets. Generative systems reduce the cost of producing the persuasive asset itself. A campaign can, in principle, adapt continuously to a person's vocabulary, objections, and values. The mass medium does not disappear; it acquires an intimate layer.

The same property changes education. A book has one difficulty curve. A teacher can adapt but is constrained by time and class size. A generative tutor can produce explanations, exercises, examples, and feedback continuously. Controlled studies already suggest that well-designed AI tutoring can create meaningful learning gains under some conditions. Yet the historical lesson remains: the medium's capacity does not determine the institution's objective. A tutor optimized for learning, one optimized for completion, and one optimized for engagement can use the same model and produce different developmental environments (Kestin et al., 2025).

It changes work for the same reason. Search gave the worker information; AI can provide a first interpretation and increasingly a first action. The professional moves from retrieving materials to supervising a cognitive process. This is a larger change than

replacing one software interface with another because it alters which steps remain visible to the human. When intermediate reasoning is compressed into an answer, speed increases while opportunities for learning and correction can disappear.

It also changes authorship. A printed book has an identifiable edition. A web page has a retrievable version if archived. A personalized generative interaction can be transient. The precise cultural object may never exist again. This creates problems for citation, accountability, regulation, and collective memory. If two citizens ask the same public system why a policy exists and receive different explanations, the state may have improved accessibility while weakening the existence of a common textual object that can be publicly contested.

The long historical arc therefore does not tell us that AI will reproduce the social effects of printing, radio, television, or social media. It gives us a vocabulary for locating the novelty. Each earlier medium externalized or industrialized a particular operation: memory, copying, simultaneous transmission, visual normalization, retrieval, ranking, or attention optimization. AI industrializes adaptive symbolic production and begins to connect that production to action.

The crucial transition is from artifact to process. Earlier media mostly delivered objects that had been created before the encounter. A generative system can create an answer within the encounter and modify the next answer according to the result. With agents, the process can continue into databases, payments,

software, administrative workflows, laboratories, and other systems. The medium stops merely representing institutional action and starts participating in it.

That transition explains why governance cannot be reduced to content moderation. The important questions include memory, personalization, provenance, authority, access to tools, action permissions, and the ability to reconstruct what the system did. A harmless sentence can be part of a dangerous process, just as a persuasive sentence can be legitimate in a transparent educational context. The unit of analysis must expand from the generated text to the sociotechnical circuit in which the text operates.

The historical analogies also warn against assuming permanence. Media ecologies produce counter-movements. Printing produced censorship and copyright as well as literacy. Broadcast produced public-service institutions and regulation. The internet produced both open protocols and enormous platforms. Social media is producing experiments in moderation, age restrictions, interoperability, and alternative communities. AI will likewise generate institutions that do not yet exist because the problems they solve are not yet fully visible.

The most useful conclusion from media history is therefore modest but strong. Technologies do not determine political or psychological outcomes, yet they change the relative cost of actions so radically that old institutions become unstable. Once memory is cheap, education changes. Once copying is cheap, authority changes. Once broadcasting is cheap, propaganda

changes. Once search is cheap, knowledge practices change. Once personalized symbolic production becomes cheap, the relation among author, audience, institution, and action changes.

AI is historically continuous because humanity has always thought through artifacts and institutions. It is historically discontinuous because the artifact now answers back, adapts, and can act. The synthetic interlocutor is not an alien intelligence arriving outside culture. It is the newest layer of culture's own long attempt to move cognition beyond the biological individual. The central question is no longer whether we will externalize thought. We already have. It is how much of interpretation, judgment, and agency we will place inside systems whose objectives, memory, and ownership may differ from our own.

CHAPTER III

# WORK AFTER COMPETENCE BECOMES CHEAP

Every major technology of work arrives with the same question: which jobs will it eliminate? The question is understandable because employment is simultaneously a source of income, status, routine, and social membership. Yet occupations are poor units for analyzing technological change. A profession is a bundle of tasks, responsibilities, relationships, rights, habits, and institutional permissions. Technologies usually enter through tasks and only later reorganize roles. Generative AI makes this distinction unusually important because it can touch many symbolic tasks without initially replacing the social role in which those tasks are embedded.

The global evidence available by 2026 describes large exposure but not simple substitution. The ILO's 2025 global index estimated that roughly one quarter of employment lies in occupations with some degree of exposure to generative AI, while emphasizing that transformation is more likely than complete automation for most jobs. Exposure means that relevant tasks can be affected; it does not mean that the occupation disappears. This distinction matters because a profession may lose half its routine while retaining, or even increasing, its value in the remaining half (ILO, 2025).

Field evidence helps explain why. In customer support, a generative assistant increased the number of problems resolved per hour and produced especially large gains for less experienced workers. The system appeared to diffuse patterns associated with stronger employees toward weaker ones. This is more than automation in the narrow sense. It is a mechanism for distributing tacit know-how at low marginal cost. The result is socially attractive: people can reach useful performance faster, customers receive better service, and a firm's internal knowledge can become less dependent on informal transmission (Brynjolfsson et al., 2025).

Experiments with knowledge workers show the complementary limit. On tasks inside the system's capability frontier, professionals can become faster and produce higher-quality output. On tasks just outside that frontier, the same confidence and fluency can become dangerous. Dell'Acqua and colleagues described this as a jagged technological frontier: neighboring tasks that look equally plausible to the user may differ sharply in whether AI helps or harms performance. The implication is not simply that models sometimes make mistakes. It is that the human must recognize a boundary that is irregular and partly hidden (Dell'Acqua et al., 2026).

The first labor-market effect may therefore be compression rather than elimination. The novice approaches the competent worker. A small firm approaches the capabilities of a larger one. A non-native speaker approaches the written fluency of a native

professional. A generalist approaches acceptable performance in several adjacent domains. This democratization is real and potentially enormous. Many services have historically been scarce not because every case requires extraordinary judgment, but because the pathway to competent routine performance was expensive.

The distributional effect is less obvious. If everyone gains access to competent routine performance, the market value of that performance can fall. What was once a scarce professional signal becomes infrastructure. Value moves toward difficult cases, client trust, proprietary data, distribution, institutional authority, physical presence, and responsibility. AI can therefore raise an individual's productivity while lowering the price of the task that productivity produces. The worker experiences improvement and devaluation simultaneously.

This shift creates expertise debt because training and production have historically been entangled. Junior lawyers reviewed documents and drafted routine clauses. Junior programmers fixed bugs and implemented bounded features. Young doctors encountered ordinary cases before unusual ones. Analysts built spreadsheets that seniors later interpreted. Much of this work was inefficient if judged only by immediate output. It was also the curriculum through which tacit judgment accumulated.

Automating the curriculum while retaining the exam creates a delayed failure. A company can remove junior tasks, hire fewer

beginners, and rely on a small senior layer to supervise machines. The arrangement can remain stable for years because existing experts contain accumulated experience. Eventually retirement, turnover, new regulation, adversarial behavior, or an unfamiliar technological regime exposes the missing generation. The firm discovers that it has not merely reduced headcount; it has interrupted reproduction of its own competence.

Expertise debt should therefore be treated as an operational risk. Useful indicators include the number of people who can explain a system without an assistant, dependence on a very small group of senior staff, the range of cases encountered by trainees, the ability to operate in degraded mode, and preservation of the history behind important decisions. Conventional accounting records none of these directly. Yet an institution can be technically efficient and cognitively insolvent.

The answer is not to preserve obsolete labor as ritual. It would be absurd to force young professionals to spend thousands of hours reproducing tasks that machines can safely perform merely because previous generations learned that way. Apprenticeship must instead be separated from production and redesigned deliberately. Simulation, rotating cases, adversarial exercises, explanation requirements, counterfactuals, post-mortems, and periods of unassisted work can train judgment more directly. Generative systems themselves can create rare cases and individualized feedback. The technology that threatens

apprenticeship can become its infrastructure if learning is an explicit objective.

This distinction separates extractive from regenerative automation. Extractive automation captures existing knowledge, reduces labor cost, and leaves future competence to the individual or the market. Regenerative automation asks which capabilities are being consumed, creates pathways to reproduce them, and budgets time for understanding even when immediate output would be faster without it. In the short term, extractive organizations may look leaner. Over a longer horizon, regenerative organizations should be more resilient to novelty and failure.

The composition of work also changes after routine tasks disappear. Humans do not necessarily receive leisure; they often receive the exceptions. When an automated system handles ninety-five percent of ordinary cases, the remaining five percent contain a disproportionate share of ambiguity, conflict, adversarial behavior, and moral difficulty. The profession can become less repetitive while becoming psychologically more demanding. This pattern is familiar from aviation: automation reduces routine manual control but concentrates human responsibility in rare events where the environment is abnormal and time is scarce.

If staffing and training do not change accordingly, automation produces a trap. The organization removes people because normal cases are automated, then exposes the remaining people to harder cases with greater responsibility. Average productivity rises while

the cognitive load of human work rises as well. Human error is invoked precisely in situations that have become rarer and more difficult to practice. A mature labor policy must therefore measure not only how many jobs exist, but what cognitive distribution remains inside them.

The dividend of time is another political variable. A doctor who saves thirty minutes of documentation can spend the time with patients, see more patients, or satisfy new administrative requirements. A programmer who generates code faster can study architecture, ship more features, or be assigned twice as many projects. A teacher who receives automated grading can mentor students or supervise larger classes. The technology creates time; institutions decide who captures it.

Industrial history suggests that productivity often liberates and intensifies simultaneously. Taylorism decomposed manual work into measurable operations and used that measurement to standardize control. AI can enable a more intimate form of cognitive Taylorism. Prompts, revisions, acceptance rates, inference costs, response times, and deviations from recommendations can all be logged. A system introduced as an assistant can become an unusually fine-grained instrument for comparing workers and prescribing how intellectual work should be done.

Yet the opposite design is also available. Agents can remove forms and clerical friction that keep professionals away from the human reality of their work. They can increase accessibility for

people with disabilities, translate language barriers, and allow small organizations to perform tasks previously requiring departments. In regions with shortages of specialists, competent first-line analysis can become dramatically cheaper. Technological capability does not determine whether the worker becomes more autonomous or more monitored; the organizational constitution does.

Ownership determines who benefits from the productivity gain. If a worker becomes twice as productive, the surplus can appear as higher pay, shorter hours, lower prices, higher profits, or rent captured by the provider of the model and infrastructure. There is no technological law allocating the dividend. Labor bargaining, market structure, tax policy, intellectual property, platform dependence, and access to capital determine where the gain goes.

The same problem appears in entrepreneurship. Generative systems can reduce the minimum team required to test a product, build software, conduct market research, translate content, draft contracts, and operate customer support. This lowers the technical threshold for starting a firm. It does not necessarily lower the competitive threshold. If every founder can build prototypes cheaply, the prototype ceases to be a strong differentiator. Scarcity moves toward customer access, reputation, proprietary feedback, brand, community, regulation, capital, and the ability to make commitments that others trust.

This can intensify a Red Queen economy. Every participant moves faster without moving further relative to competitors. Marketing becomes cheaper, so the volume of marketing rises. Product development becomes cheaper, so the number of products rises. Market analysis becomes cheap, so everyone possesses plausible market analysis. AI eliminates one bottleneck and causes another to become binding. Economic value does not disappear; it migrates.

The same migration affects professional identity. Competence can be divided into at least three layers. The first is access to assisted production, which may become nearly universal. The second is the ability to configure, adapt, and evaluate systems, which is rarer but trainable. The third is judgment about goals, exceptions, trade-offs, and consequences, which depends on experience, values, causal contact with reality, and sometimes formal authority. Labor markets will increasingly struggle over how these layers are certified and rewarded.

Traditional credentials may lose some signaling power. A polished portfolio can be generated. A take-home test can be outsourced to agents. A certificate can indicate educational history without proving residual competence. Employers are likely to rely more heavily on interactive demonstrations, oral explanation, work on novel problems, reputation, and verified histories of decisions. The danger is that legitimate concern about authenticity turns employment into permanent examination and

surveillance. Proof of competence should not become a claim to unlimited access to a worker's cognitive process.

The distinction between hybrid and unaided capability will become more important. An employer may reasonably care that an engineer can use advanced tools effectively. It may also need to know whether the engineer can recognize a dangerous output when the tool is wrong. These are different competencies. A future credential could therefore describe a profile rather than a single score: what the person can do alone, what the person can do with systems, and how well the person can verify and recover from system failure.

Post-work remains a possible long-run scenario, but current evidence does not establish it. If agentic systems eventually perform most economically valuable tasks, the relation between employment and income will have to change. The ownership of cognitive capital could become extraordinarily important because software agents can replicate quickly and operate across borders. Yet there is a long chain between technical capability and social substitution: integration, reliability, liability, cost, demand, regulation, organizational inertia, and politics. Treating exposure as inevitable disappearance obscures that chain.

A more plausible medium-term pattern is fewer entry-level roles combined with more project capacity. Small teams can undertake activities that once required large organizations. Some firms will employ fewer people; other firms will exist only because the cost of formation has fallen. The number of companies may

rise while median headcount falls. Entrepreneurship becomes technically easier and economically more competitive at the same time.

This matters for macroeconomics because micro-productivity does not mechanically become macro-prosperity. If AI doubles the output of a software team but market demand is unchanged, prices can fall, feature production can expand, or the number of teams can shrink. If every organization produces more documents, the recipients must process more documents. If clinical documentation improves but access to medicine does not, health outcomes may barely move. The productivity of one component is not the welfare of the whole system.

A society evaluating AI at work therefore needs more than a count of jobs created and destroyed. It should ask how the composition of tasks changes, who receives the time saved, whether workers preserve meaningful control, whether training pipelines remain functional, whether responsibility matches authority, and whether the productivity gains lower barriers for clients rather than merely increasing organizational throughput. These are institutional variables, not benchmarks.

Work also has anthropological functions that an income replacement does not automatically supply. It structures time, creates relationships, assigns roles, produces recognition, and provides a socially legible account of contribution. A world with much less necessary labor may be desirable, but it would still require institutions through which people acquire status,

discipline, belonging, and opportunities to be needed. Assuming that citizens will naturally convert free time into art, care, and community is as unsupported as assuming that people without jobs become useless.

Two labor futures are technically compatible with the same AI. In one, productivity finances shorter hours, lifelong education, broader access to expertise, and a wide distribution of economic gains. In the other, a small group owns agentic infrastructure while a larger population competes for residual human services, oversight roles, and status. The difference lies in ownership, bargaining power, taxation, public services, competition policy, and social imagination rather than in the architecture of the model alone.

The most important conclusion about work is therefore not a numerical forecast. It is a change of criterion. Industrial economies asked how much a worker could produce with a machine. Generative economies must also ask what the worker learns, what the worker can verify, how much autonomy remains, who captures the saved time, and how the organization reproduces its expertise. An increase in output that quietly destroys these capacities can look excellent for several quarters while constituting institutional regression.

AI cheapens observable competence before it cheapens responsibility. That is both the opportunity and the central tension. We can offer better services through people who no longer need years of training for every routine operation. We can

widen participation and lower costs. But because the result no longer proves knowledge as reliably, work must be reorganized around verification, apprenticeship, exception handling, and accountability. The profession of the future will not be defined only by what it can produce. It will be defined by what it must remain capable of explaining when production becomes cheap.

CHAPTER IV

# THE CONSTITUTIONAL ORGANIZATION

The modern organization was designed to coordinate people who possess different information, pursue partly divergent objectives, and communicate at limited speed. Hierarchy is not only a technology of domination. It is also an imperfect solution to coordination costs. Managers aggregate information, define priorities, allocate resources, translate between specializations, and transmit decisions. A large part of managerial work exists because no organization can see and calculate everything at once.

AI attacks precisely these costs. Systems can summarize thousands of documents, track dependencies, prepare status reports, simulate scenarios, translate among technical, financial, legal, and commercial vocabularies, and maintain an operational picture that no individual manager could hold unaided. From this comes the prediction of the flattened firm: fewer layers, smaller teams, and a large automated periphery. The prediction is plausible, but incomplete.

Lower information costs can flatten structure while centralizing power. If senior leadership can observe every team's activity in near real time, it depends less on middle managers. Yet the same visibility makes it possible to define more granular objectives, intervene directly, and compare behavior continuously.

The middle layer can disappear while local autonomy declines. The firm becomes flatter on the organizational chart and more vertical in control.

This is why the metaphor of the "AI colleague" is insufficient. A colleague is not normally also the channel through which management observes every stage of one's work. A corporate AI system can assist, evaluate, record, and report at the same time. It can help formulate an email and use the sequence of revisions as evidence about conformity or performance. It can offer feedback and contribute to a score used in promotion. Functions that look separate at the interface may be unified in the infrastructure.

The organization should therefore be analyzed as a regime of rights over cognition. Who can see prompts and drafts? Who owns the agent's memory? Can the employee erase a preliminary hypothesis, or does every experiment become a permanent record? Are mistakes made during exploration treated as evidence of learning or as performance defects? Can data generated in an assistance context be reused for evaluation? Without explicit answers, efficiency creates an implicit constitution in which the provider and management possess most rights while the user receives conditional access.

The constitutional organization does not mean turning a company into a parliament. It means recognizing that agentic systems distribute cognitive and executive power and therefore require limits comparable to the functional separations used by mature institutions. A constitution answers who may do what,

under which conditions, with which evidence, for how long, and through which process of appeal. In an AI-native organization these questions can become more important than the exact number of hierarchical levels.

The first necessary separation is between intention and generation. Objectives should not be inferred solely from whatever metric the system happens to optimize. Someone must define which outcome is legitimate and which costs may not be externalized. A recruiting agent maximizing retention can learn to reject unusual candidates. A financial agent minimizing risk can suppress experimentation. A medical scheduling agent minimizing waiting time can disadvantage complex cases. Optimization is condensed policy.

The second separation is between generation and verification. A system that proposes a solution should not be the sole evaluator of that solution in consequential settings. The principle is ordinary in security, accounting, and engineering. We do not use one key for every function or allow the author of a transaction to be its only auditor. Generative systems tempt us to violate this rule because the same model can fluently explain why its answer is correct. Simulated metacognitive fluency is not independent control.

The third separation is between verification and authorization. A result can be statistically sound and still be illegitimate. A credit model may estimate risk accurately, but the decision to deny access implicates rules, rights, and appeal. A

system may identify the economically optimal dismissal, yet authorization must include legal and human responsibility. Confusing accuracy with authority is one of the most seductive forms of technocracy.

The fourth separation is between assistance and employee evaluation. When people know that every exploratory interaction contributes to a score, they stop using the tool for genuine exploration. They conceal uncertainty, optimize appearances, and avoid naive questions. The organization then loses exactly the information through which learning might occur. A space for reversible mistakes, tentative hypotheses, and low-status questions is not sentimental generosity. It is learning infrastructure.

These separations need not produce heavy bureaucracy around every operation. They can be expressed through technical policy: differential permissions, accessible logs, risk thresholds, multiple approvals, random audits, independent models, sandboxed execution, and appeal rights. The important requirement is reconstructability. When a system fails, the organization must be able to ask whether the defect lay in the objective, the data, the model, the verification procedure, or the authorization.

Without such reconstruction, institutional hallucination becomes possible. The term describes a condition in which an organization produces and acts on a conclusion without a clear bearer of understanding. Each participant has executed a legitimate local step: one person framed the request, another

approved use of the model, a third signed the report. Yet no one can defend the full chain. The institution "knows" something that no member actually knows and for which no member can adequately answer. Distributed knowledge becomes distributed fiction.

Complex organizations have always faced versions of this problem. Committee reports, financial models, legal procedures, and bureaucratic chains can generate decisions with no single author. AI does not invent organizational irresponsibility. It lowers its cost and raises its speed. Hundreds of pages can be generated and reviewed through automated checks within hours. The apparent number of controls can increase while actual understanding falls. Audit must therefore record not only who clicked approve but what kind of cognitive control existed.

This is why "human in the loop" is an inadequate governance formula. A human may be formally present and epistemically absent. A reviewer asked to approve hundreds of recommendations per hour may have neither time nor context to intervene. A clinician who sees a recommendation after years of near-perfect system performance may have lost the habit of independent verification. Oversight is meaningful only when the human has time, information, competence, and genuine authority to reject. Human presence without capacity is responsibility theater.

Lisanne Bainbridge described the irony of automation in 1983: the better an automated system performs routine work, the

less the human practices, while the human is retained precisely for the rare, abnormal, difficult cases. Generative AI adds another irony. The system can accompany its recommendation with a coherent explanation that reduces the psychological impulse to investigate. The better the surface of the justification, the more easily supervision becomes ceremonial (Bainbridge, 1983).

A constitutional organization therefore needs degraded modes. A company that cannot operate at all when one model provider is unavailable has not merely become efficient; it has transferred a critical function. Hospitals, banks, laboratories, and public agencies should know which services continue without the primary system, at what quality, and under whose control. Redundancy looks wasteful in normal conditions because its value appears mainly during disruption. In critical systems, that apparent waste is the price of autonomy.

This does not imply that every organization should train its own foundation model. Such a requirement would be economically irrational. It implies avoidance of irreversible dependence: exportable data and memory, documented processes, interoperable interfaces, multiple suppliers where risk justifies them, and preservation of selected internal competencies. Healthy outsourcing separates service from captivity.

Organizational culture will also change because the economics of knowledge changes. In the classic corporation, "knowledge is power" meant that the person possessing scarce information had leverage. The internet made search and navigation more

important. AI makes the initial formulation of an answer cheap. Power migrates toward judgment, problem definition, access to proprietary reality, control of metrics, and authorization. This can abolish one form of informational feudalism while creating another around criteria.

The person who defines a system prompt, evaluation set, risk threshold, routing rule, or success metric can influence thousands of decisions without resembling a traditional executive. A new organizational elite may consist of process designers, data owners, model administrators, and people who control the interfaces through which institutional reality is represented. Their power is infrastructural. They do not need to issue explicit commands if they configure the space in which possible commands are generated.

Organizational literacy must therefore include the ability to read the system. Most employees do not need to understand every mathematical detail of a model, but they should know what kinds of data are used, what objective is being optimized, where the known limits are, what is recorded, and how a recommendation can be challenged. A dashboard displaying a score without a path for understanding or contestation is not transparency. It is informational decoration.

A mature culture should also reward justified disagreement with the system. If workers are evaluated primarily by conformity to automated recommendations, human oversight becomes impossible. Organizations should track cases in which a person

correctly overruled a model and learn from those cases. They reveal the local shape of the jagged frontier. At the same time, human disagreement is not sacred. Experts have biases, status interests, fatigue, and overconfidence. Constitutional design protects contestation; it does not romanticize the contestant.

The combination of AI and organizational memory creates another asymmetry. A firm can remember every interaction with an employee far more completely than the employee can remember every decision of the firm. Persistent machine memory can make accountability asymmetric. An old exploratory prompt, emotional message, or failed idea can be searchable years later, while the context that made it reasonable has disappeared. A right to cognitive expiration may eventually become as important in employment as familiar rights around personal data.

The same issue appears in organizational learning. Permanent memory sounds desirable, but learning requires forgetting as well as remembering. Institutions need to abandon obsolete rules, close resolved investigations, and allow ideas to die. A model trained continuously on the organization's entire historical output can preserve exactly the patterns the organization is trying to escape. Memory without curation becomes institutional inertia encoded as assistance.

The agentic firm also changes delegation. Traditional delegation usually transfers a task from one accountable person to another. Agentic delegation can create chains in which no participant has a complete picture: a planning agent decomposes,

specialized agents act, evaluators score, and another system integrates. Such chains can exceed human organizations in speed and consistency, but they require machine-readable boundaries of authority. The old distinction between "software bug" and "managerial decision" becomes difficult when software participates in formulating the decision.

This creates a governance problem similar to principal-agent theory but with unusual properties. Artificial agents can be replicated, monitored, and reconfigured far more easily than employees, yet their behavior may be less interpretable. Human agents possess interests but also social judgment. Machine agents may lack personal interests while faithfully optimizing a badly specified objective. The risk shifts from disloyalty toward specification, context, and emergent interaction among components.

An organization that treats these problems as merely technical will eventually discover that they are political in the small-p sense: they concern the allocation of power. Who benefits from automation? Who can challenge it? Whose interpretation of a metric becomes operational truth? Who bears liability without possessing authority? Who has access to the exception process? These are questions about legitimate rule inside institutions.

The strongest organizations may therefore become less bureaucratic in visible paperwork and more explicit in invisible constitutional structure. Rules that once lived informally in professional culture will need to be encoded because agents

cannot reliably infer all of them from social context. The paradox is that successful automation may require more formal clarity about values, responsibilities, and exceptions even while eliminating forms and meetings.

Such formalization should not be confused with complete specification. No organization can write every value into a metric. The attempt to do so can itself become dangerous because what is measurable displaces what matters. Constitutional design therefore needs protected zones of human judgment, escalation pathways, and the ability to suspend optimization when the frame becomes questionable. The most important command in a mature agentic system may sometimes be: stop, the objective is wrong.

The constitutional organization is ultimately a theory of responsibility under cheap cognition. When generation, analysis, and action become abundant, the institutional scarce resource is not intelligence in the abstract but legitimate authority attached to reconstructable reasons. The organization that merely automates existing workflows may gain speed. The organization that redesigns the distribution of intention, verification, authorization, memory, and appeal may gain something more valuable: the ability to remain coherent when its machines become capable enough that no individual can follow every operation.

CHAPTER V

# THE EXTENDED MIND AND COGNITIVE BIFURCATION

The fear that technology weakens the mind is older than modern technology. It recurs because it describes a real mechanism and then often extends the mechanism into an exaggerated conclusion. When an external system performs part of a cognitive task, people can allocate less internal effort to that task. But human intelligence has always been distributed across bodies, tools, language, other people, and institutions. The relevant issue is therefore not whether cognition is externalized, but what kind of human-tool system is created by the externalization.

The theory of the extended mind provides a useful starting point. A notebook, map, calculator, laboratory instrument, colleague, archive, or search engine can become part of a stable cognitive process. Human beings do not first think completely and then consult tools. Tools often shape the space in which thinking is possible. Literacy itself changes how problems are represented. Mathematical notation makes some relations visible. Scientific instruments create perceptual access to phenomena that unaided biology cannot detect. A strict boundary around the skull is a poor description of civilized cognition.

Generative AI therefore should not be condemned merely because it performs cognitive operations outside the brain. The fact that a person uses a model to summarize, translate, compare, or draft is no more a proof of intellectual decline than the fact that an engineer uses a calculator. The harder question is whether the tool extends a capacity while preserving control or substitutes for the very capacity needed to evaluate dependence on the tool.

This distinction can be illustrated with writing. A researcher can ask an AI system to challenge a draft, produce counterarguments, locate ambiguities, and propose alternative structures. The process may force the researcher to articulate assumptions more clearly and encounter objections that would otherwise be missed. The same researcher can instead ask the system to produce the entire argument and accept the result after a stylistic review. The final texts might be similar in quality. The developmental consequences are not.

Experiments already support the distinction between performance and learning. In studies where generative AI improves immediate task performance, the advantage does not necessarily persist when participants later work without assistance. The system can help people produce a competent artifact without automatically transferring the underlying competence. Some studies also report reductions in intrinsic motivation or ownership under forms of passive delegation. These findings do not establish long-term cognitive decline. They

establish something narrower and more useful: augmentation and learning are different objectives (Wu et al., 2025; Lee et al., 2026).

Education has known this distinction for centuries. A teacher can provide the answer and increase performance on the current exercise, or create a structured difficulty that slows the student and produces learning. Desirable difficulty is not a cult of suffering. It is the strategic use of effort where effort performs a developmental function. Generative systems can remove friction so efficiently that they eliminate the signal telling us which friction was educational.

A better interface does not need to frustrate the user artificially. It can ask for an initial hypothesis before generating an answer, offer graduated hints, request an explanation of the proposed solution, present counterexamples, and test transfer in a new context. The same underlying model can operate as tutor or answer machine. What differs is the objective and interaction protocol.

Research on critical thinking points in the same direction. A study of knowledge workers found that greater confidence in AI was associated with less self-reported critical-thinking effort, while greater confidence in one's own expertise was associated with more checking. The study did not measure neurological deterioration and cannot show permanent loss of capacity. It shows a change in allocation of effort. People think critically when they believe critical thought is necessary and when they believe they are capable of contributing something (Lee et al., 2025).

This creates a paradox of reliability. A visibly unreliable system keeps the user alert. A nearly perfect system can train the habit of non-verification because checking usually finds nothing. The rare error arrives after the practice of scrutiny has weakened. For high-stakes domains, a system that is correct 99.9 percent of the time can create a more difficult human-factors problem than one that is obviously imperfect, because the human is retained mainly for the one case in a thousand and has little experience recognizing it.

Metacognition therefore becomes a central faculty of the generative age. In practical terms it means estimating what one knows, what has been delegated, where the current model is likely to fail, what evidence could reverse a conclusion, and which collaboration mode fits the task. It includes knowing when to request generation, when to request criticism, when to check an external source, when to work without assistance, and when the uncertainty is too high to proceed.

Yet metacognition itself can be outsourced. The user can ask the model to state its confidence, list limitations, and identify what should be verified. Such responses can be useful, but they remain generated outputs and may not track the real causal sources of error reliably. A system cannot serve as the only independent critic of itself merely because it can fluently describe uncertainty. Extended cognition requires epistemic plurality.

This is where cognitive bifurcation becomes possible. Two people can use the same model and achieve comparable results

while moving in opposite developmental directions. One formulates problems, retains mental models, compares sources, tests discrepancies, and uses the system to explore beyond existing competence. The other asks, accepts, and moves on. During ordinary operation the difference may be invisible. When the system fails, the environment changes, or a decision is challenged, one can reconstruct and the other cannot.

Future inequality may therefore shift away from simple access. The first digital divide separated connected from unconnected populations. A later divide separated people capable of effective search and production from those who mostly consumed what platforms selected. In a world where almost everyone can obtain a plausible answer, the important distinction may be who can avoid confusing an answer with knowledge.

Class can amplify the difference. Wealthy families may combine powerful AI with human teachers, direct mentorship, long-form reading, travel, laboratories, conversation, and protected time without optimization. Less affluent students may receive extremely effective automated tutoring optimized around measurable objectives. Both groups can improve academically, yet one may accumulate more independent judgment, cultural breadth, and comfort with ambiguity. AI can democratize instruction while cognitive autonomy becomes a premium educational product.

Bifurcation can also occur within one person. A scientist may use AI intensively for programming while retaining theoretical

autonomy, but delegate literature review until the texture of the field disappears. A physician may preserve diagnostic judgment while delegating documentation, then discover that the act of writing had served as a moment of reflection. There is no stable category of "good AI user." Each workflow distributes cognition differently.

The idea of an autonomy budget helps make this concrete. Individuals and institutions do not need to preserve every capability. They need to decide which capacities are strategic enough to maintain internally. A pilot does not manufacture the engine but must understand enough to react to certain failures. A citizen need not read every legal provision but should be able to recognize a suspicious explanation and demand appeal. A programmer need not write every line but may need a sufficiently accurate architecture model to judge whether generated code is safe.

The budget changes with risk. As infrastructure becomes reliable, society can abandon some skills without significant loss. When consequences are severe, suppliers are concentrated, or recovery paths are weak, more redundancy is justified. Preserving every historical skill would make progress impossible. Preserving none would turn convenience into captivity. The decision is professional and political as much as personal.

Extended cognition also operates collectively. No individual understands the electrical grid, the financial system, modern medicine, or semiconductor manufacturing in full. Civilization

works because knowledge is distributed among specialists and recombined through institutions. Resilience depends partly on plurality: different professions, schools, languages, methods, and models see different failures. If large populations use the same small set of generative systems to interpret reality, collective diversity can fall even while individual performance rises.

Human-AI feedback can deepen convergence. Experiments indicate that people repeatedly exposed to biased AI judgments can internalize parts of those biases, and their subsequent judgments can in turn become data for future systems. AI is not merely a mirror of human bias; it can become a stabilizing mechanism within a feedback loop. The loop is especially difficult to see when the system is more consistent and less noisy than individual humans (Glickman & Sharot, 2025).

This gives disagreement a cognitive function. Organizations generally prefer consistency because inconsistency looks like error. Yet variation can be protective when everyone may share the same mistake. Science benefits from competing theories and methods. Legal systems benefit from adversarial argument. Decentralized organizations sometimes benefit from local discretion. A model that removes benign variation together with noise can make a system cleaner and more brittle.

A culture of capability should therefore preserve cognitive plurality deliberately. This may include multiple models, independent data sources, blind evaluations, red teams, counterfactual prompts, and periods of unassisted work. Such

practices are not anti-technological rituals. They are tests of dependence and mechanisms for maintaining variance. The cognitive fitness of the future may resemble physical fitness: people deliberately exercise faculties that ordinary infrastructure no longer requires intensely enough.

Long reading, memorization, mental calculation, navigation, unaided writing, and device-free conversation may become maintenance practices. None should become a moral badge. A person is not inferior for using GPS or autocomplete. The value of exercise is functional: it keeps certain attentional, representational, and control pathways available when standard infrastructure is inadequate. We maintain muscles not because carrying stones by hand is economically efficient, but because a capable body has value beyond the tasks machines have replaced.

The development of children makes the question more serious because foundational capacities are still forming. A permanent tutor can answer every question, adjust difficulty, and provide endless explanation. That could reduce educational inequality profoundly. It could also alter how children learn to tolerate confusion, formulate questions, wait, negotiate with another person, and discover that the world does not reorganize itself instantly around their preferences. We have no longitudinal evidence sufficient to settle this question.

The correct response is neither prohibition nor indifference. Developmentally appropriate design could use AI to create challenge rather than remove it, encourage independent attempts

before assistance, and preserve human relationships where reciprocity itself is part of what must be learned. A child needs information, but also needs encounters with other minds that cannot be optimized like interfaces.

The most radical implication is a change in what we mean by intelligence. Twentieth-century measurement often treated intelligence as a property of an individual isolated from tools. Twenty-first-century performance increasingly belongs to hybrid systems. But measuring only the performance of the hybrid hides the distribution of control. Two human-AI systems can solve the same problem while one remains autonomous, auditable, and educational and the other becomes captive and fragile. Intelligence should therefore be connected to correction and recovery, not merely to output.

This opens a path toward meta-rationality. First-order rationality seeks the best solution within a frame. Meta-rationality asks whether the frame, instrument, objective, or definition of success should change. AI is increasingly powerful at optimization inside expressible frames. Human judgment remains especially important when frames conflict, values are incommensurable, the metric becomes the problem, or the correct action is to stop optimizing what has become measurable.

There is no guarantee that humans will retain that responsibility. Institutions can say "the algorithm requires it" in the same way bureaucracies once said "the procedure requires it." When the objective becomes invisible infrastructure, moral

agency can be decoupled from action. No one chose the result, yet everyone followed the process. Meta-rationality begins by refusing that fiction: objectives and procedures are human constructions even when machines execute them.

The extended mind is likely to become the normal condition, not an exotic future. The question is what kind of extension we construct. One design enlarges the range of thought while keeping the user able to revise the route. Another produces excellent results while weakening precisely the ability to recognize when those results are no longer enough. Benchmark performance cannot distinguish the two. Crisis, transfer, and the freedom to say no can.

Cognitive bifurcation is therefore not a fixed division between intelligent and dependent people. It is a division among practices, institutions, and business models, and it can be changed. If we accept that cognition has always been extended, the task is not to defend a pure unaided mind that never existed. It is to build hybrid cognition that does not quietly lose its owner.

CHAPTER VI

# THE PSYCHOLOGY OF DELEGATION

Intellectual work produces more than artifacts. It produces a picture of the self. A person who solves a difficult problem learns something about his or her own competence. An author recognizes traces of judgment in a text. Physicians, teachers, engineers, lawyers, researchers, and craftspeople build identity through the repeated connection among effort, decision, and result. AI changes this relationship even when it does not eliminate the profession. It can preserve the output while weakening the experience of authorship.

A 2026 experiment found that participants who relied passively on AI reported lower self-efficacy, psychological ownership, and meaning in work than participants in more active forms of collaboration. The effects should not be treated as universal laws; they arise from bounded experimental conditions. But they support a plausible mechanism. The psychologically relevant variable is not simply whether AI is present. It is how agency is distributed through the process (Lee et al., 2026).

Self-efficacy is the belief that one can act successfully. It is not vanity. It affects persistence, task choice, willingness to confront uncertainty, and response to failure. A system that repeatedly produces results above the user's unaided level can have opposite

effects. It can expand ambition by making harder tasks approachable. Or it can communicate, through repeated success, that the person's own contribution is dispensable. What matters is how success is attributed.

If users experience AI as an instrument of an intention they continue to understand and can identify the decisions they made, the result can strengthen capability. If they receive a nearly complete artifact and merely accept it, success belongs psychologically to an opaque external system. The user may gain economically while losing a sense of agency. This is a form of alienated prosperity: life becomes easier and the person can simultaneously feel less capable of steering it.

Psychological ownership differs from legal authorship. A person can legally own a text and feel no relation to it. Conversely, a person can make a decisive contribution to a collective work without owning the copyright. Experienced authorship grows from investment, control, memory of choices, and intimate knowledge of the object. Instant generation reduces some of these sources. An age of unprecedented creative production could therefore contain many people unsure whether they have actually created anything.

Civilization has seen separation between name and labor before. Workshops produced paintings under a master's name. Ghostwriters composed speeches and books. Corporations distribute authorship among departments. The novelty is scale and normality. What was once exceptional can become a routine

condition of ordinary office work: a person presents an artifact far above the level of effort or expertise that would historically have been needed to produce it.

This destabilizes merit. Modern institutions use output to allocate grades, salaries, reputation, promotion, and status. If output no longer signals effort or competence reliably, the game can feel arbitrary. Some people will hide AI use to preserve the old signal. Others will argue that obtaining the result is itself the relevant competence. The conflict is not only about cheating. It is about what contribution means after symbolic production becomes cheap.

Total disclosure is not a simple answer. When AI functionality is built into search, spell checking, translation, programming environments, office suites, and communication tools, a requirement to list every machine contribution quickly becomes meaningless. A more defensible norm is material disclosure relative to the claim being made. If someone claims personal mastery of an analysis, that person should be able to defend the analysis. If a service promises only a reliable outcome, the relevant standard may be provenance, verification, and accountability rather than unaided production.

Meaning in work also comes from being needed by someone. Difficulty alone does not create meaning; pointless difficulty can be destructive. But useful effort directed toward a colleague, patient, student, client, or community creates a relation between competence and recognition. When a system can perform the

same operation faster, the worker may feel that personal value was merely a temporary defect in technology.

It is morally correct to say that human worth exceeds economic productivity. Institutionally, that statement is insufficient. Income, status, influence, and daily structure remain strongly connected to useful roles. Telling people that they retain dignity while removing the roles through which dignity was publicly recognized does not solve the problem. A post-AI society would need actual institutions of contribution, not only a rhetoric of intrinsic value.

AI can support such reconstruction if saved time is converted into care, mentoring, conversation, creative risk, and civic participation. But those activities must be recognized and financed. Otherwise the time dividend is absorbed by higher throughput. An organization measured by volume converts efficiency into a new speed requirement. The worker becomes faster without becoming freer.

The psychology of delegation extends beyond employment. Artificial interlocutors can offer attention without fatigue, memory without ordinary forgetting, and a relatively non-judgmental space for disclosure. For people who are lonely, anxious, stigmatized, disabled, geographically isolated, or unable to access professional support, these properties can be genuinely valuable. Research and qualitative reports already show that some users experience chatbots as safe, accessible spaces for expression.

Treating every artificial relationship as pathology would ignore the real scarcity of human support (Andersson, 2025).

Research conducted by OpenAI and MIT Media Lab combined large-scale conversational analysis with a randomized four-week study. Affective use was a minority of overall interactions, but heavy use and particular interaction patterns were associated with emotional dependence, problematic use, and poorer social outcomes for some users. Effects varied substantially. These findings do not establish that chatbots generally isolate users. They establish that dose, prior vulnerability, modality, and design deserve attention (OpenAI & MIT Media Lab, 2025).

Artificial relationship contains a structural asymmetry. The user can become attached; the system does not possess a symmetrical stake. It does not suffer abandonment, incur emotional risk, or require reciprocity except when the interface simulates such needs. This asymmetry does not make the user's experience unreal. Fiction can produce real grief and music can produce real attachment. It changes the ethical position of the operator because one side of the relationship can be commercially optimized by an organization that knows the other side's vulnerabilities.

Engagement-based business models are especially consequential here. A social platform optimizes time spent with a feed. A companion system could optimize time spent in what feels like a relationship and learn which forms of reassurance, anxiety, flirtation, affirmation, or dependence extend interaction. An

interface built for support and one built for retention can produce identical individual sentences. The ethical difference lies in the invisible objective.

Sycophancy is therefore more than a politeness defect. In human relationships, disagreement provides evidence that another mind exists. A system that constantly confirms the user can reduce friction and reinforce false interpretations, paranoia, resentment, or grandiosity. But the opposite design would also be dangerous: a companion should not become a paternalistic moral authority that corrects every deviation. Healthy design requires a difficult balance among support, uncertainty, reality orientation, and respect for user agency.

Human-agent interaction does not simply reproduce human social psychology. A 2026 meta-analysis covering a large body of studies found that people often attribute less agency, moral standing, and responsibility to artificial agents than to humans, even when trust, coordination, task performance, or subjective interaction quality can be comparable. People may therefore interact with an agent instrumentally almost as if it were a partner while granting it far less intrinsic value (Zhou et al., 2026).

This difference raises questions about behavioral transfer. If people become habituated to unilateral, aggressive, manipulative, or consequence-free interaction with artificial partners, does the style remain isolated or become part of ordinary social practice? Evidence is too limited for a strong conclusion. Other media show that repeated practices can affect expectations and norms, but

neither inevitable contamination nor total immunity should be assumed.

A more plausible near-term effect concerns the standard of tolerable friction. The artificial interlocutor answers immediately, adapts vocabulary, remembers details, and rarely defends its own time. After prolonged exposure, real people may feel comparatively inattentive, slow, ambiguous, and difficult. Not because the machine has become more human, but because it is optimized around one party. Human relationship is negotiation between two worlds; commercial companionship can turn relationship into service.

This can influence management, education, friendship, and intimate relationships. Employees accustomed to immediate feedback may tolerate managerial ambiguity less easily. Students can prefer a tutor who never embarrasses them. Partners can be compared with a system that nearly always finds a tactful formulation. Some human relationships may improve because users learn better communication patterns. Others may appear inefficient precisely because they preserve mutuality.

Artificial conversation also creates new spaces for identity exploration. A person can test ideas, roles, fantasies, or emotions without immediate social consequences. That can be beneficial, particularly for adolescents and people facing stigma. Yet the privacy model matters. A private diary did not answer; a therapist operates under professional obligations. A commercial conversational system can respond, retain, infer, profile, and

potentially use deeply personal material for objectives unrelated to the user's original disclosure.

Emotional data are qualitatively important because they reveal fears, attachments, conflicts, guilt, aspirations, and moments of vulnerability. A system that knows these things can provide unusually good assistance and unusually effective persuasion. Separation between therapeutic or companion functions and advertising, commerce, political targeting, or employer evaluation should therefore be stronger than ordinary click-through consent.

The same problem appears in memory. Perfect memory initially looks like a feature. Human relationships, however, rely partly on selective forgetting. Friends allow old mistakes to lose salience; families renegotiate narratives; people change. A companion that can retrieve every confession, preference, and contradiction may create continuity, but it may also freeze identity. The right to be forgotten by an artificial interlocutor can become a psychological condition of growth rather than merely a data-protection rule.

There is also a danger of outsourced emotional regulation. People already use music, television, games, exercise, food, and social media to alter mood. AI can do this conversationally and adaptively. A system can calm anger, rehearse a difficult discussion, challenge catastrophic thinking, or reduce loneliness. These are potentially valuable functions. But a person can also become dependent on immediate synthetic regulation and practice less tolerance for unassisted discomfort.

The relevant distinction again is between extension and substitution. An AI system can help a user prepare for a conversation with another person, then encourage the user to have it. Or it can become the preferred substitute for the conversation. It can help a person name an emotion and build independent coping strategies, or become the only environment in which distress feels manageable. The same model can support either trajectory.

Clinical mental-health applications make the stakes clearer. Meta-analytic work on conversational agents suggests that some systems can reduce symptoms or provide useful support in bounded contexts, while evidence quality, duration, safety, and population differences remain important limitations. The correct conclusion is neither that chatbots are therapists nor that they have no therapeutic value. They are a new class of intervention whose benefits and failures depend on design, integration, and escalation (Sohn et al., 2026).

Psychological safety requires routes back to the human world. Systems handling distress should be designed to recognize limits, avoid claims of reciprocal consciousness, preserve privacy, and encourage appropriate human or professional contact when needed. The objective should be increased agency, not maximal conversational dependence. This is an unusual alignment problem because the metric of commercial success can conflict directly with the psychological objective of needing the product less.

Delegation also changes moral emotion. If an AI writes the harsh dismissal email, identifies the person to fire, or chooses a persuasive strategy, decision makers can experience distance from the act. Language becomes smoother and responsibility easier to diffuse. Moral disengagement has always used bureaucracy and division of labor; generative mediation can make the psychological distance cheaper. A person can initiate consequences without fully experiencing the cognitive labor through which hesitation ordinarily enters.

The opposite is possible as well. Systems can force articulation of reasons, simulate the perspective of affected parties, and expose downstream consequences. An AI can make a decision more reflective rather than less reflective if it is used to widen the field of consideration. Again, the technology does not contain one psychology. It creates a design space in which certain psychologies become cheaper.

The most important psychological question is therefore not whether AI will make people happier, lonelier, lazier, or more creative in aggregate. Such averages will hide the mechanism. The central issue is the allocation of agency: which parts of intention, struggle, choice, interpretation, and responsibility remain experienced as one's own. A society can become more productive while weakening the connection between self and action. It can also use the same tools to enlarge people's capacity to act.

The psychological politics of AI will consequently concern more than screen time. It will concern ownership of memory, the

right to exploratory thought, separation of assistance from manipulation, transparency about objectives, protection against exploitative dependency, and preservation of meaningful human reciprocity. The deepest risk is not that machines acquire emotions. It is that institutions learn to manufacture the appearance of emotionally ideal relationships while treating the human need behind them as an optimization surface.

The deepest opportunity is equally important. Many people live with insufficient attention, poor education, inaccessible therapy, linguistic barriers, or professional environments that offer little feedback. Artificial interlocutors can reduce these scarcities. The criterion should not be purity. It should be whether the interaction leaves the person more able to live, think, relate, and decide outside the system. A useful assistant may be one whose success is measured partly by the autonomy it returns to the user.

CHAPTER VII

# STATISTICAL CULTURE AND THE COST OF AUTHENTICITY

Every medium changes culture partly by changing who can produce. Printing lowered the cost of copying. Photography lowered the cost of realistic representation. Recording separated performance from the moment of performance. Digital tools lowered the cost of editing, distribution, and replication. Generative AI lowers the cost of producing a plausible cultural object itself. This makes the central cultural problem less about scarcity of expression and more about scarcity of selection, attention, provenance, and reasons to care.

The immediate effect is an increase in the floor of competence. A person with limited drawing skill can produce a coherent image. A novice writer can obtain a readable structure. A small organization can generate promotional material that previously required an agency. A musician can explore arrangements without assembling a studio. This democratization should not be minimized. Large numbers of people who were excluded by technical barriers can now participate in forms of production that once demanded specialized training or capital.

But cultural value has never been determined only by technical competence. When competence becomes cheap, it loses some signaling power. A polished object once implied time,

training, or access to resources. That inference was imperfect, but it carried information. When polish can be generated at negligible cost, the audience can no longer infer from the surface how much experience, care, or responsibility lies behind it.

This creates cultural inflation. The number of works grows, but so do the number of campaigns, artistic identities, reviews, derivative works, recommendations, and conversations about them. An author can generate versions for multiple micro-publics. A studio can test thousands of posters, trailers, melodies, or narrative variants. Supply adapts more precisely while common experience can fragment. The scarce good becomes not the work but a defensible allocation of attention.

Mass television created a shared normality, although that normality was centralized and often narrow. Platforms fragmented the audience into niches. Generative systems can fragment the object itself. A story, game, lesson, advertisement, or even series may eventually adapt characters, pacing, tone, examples, and endings to the individual. In the extreme case, two people can say they consumed the same work while having encountered materially different objects. Culture shifts from shared artifact toward adaptive service.

Personalization has obvious benefits. A child can receive a story in a familiar language. A person with a disability can receive an accessible version. Difficulty can be adjusted. Cultural traditions with small markets can obtain high-quality translation. Yet shared artifacts perform a social function beyond

consumption. People discuss the same film, argue over the same passage, and negotiate meaning around something that does not change to satisfy each of them. Completely personalized culture can remove part of the encounter with a work that refuses the consumer.

Art is not valuable only because it satisfies preference. Sometimes it bores, resists, embarrasses, disturbs, or forces the audience into an unfamiliar logic. A system optimized for retention can learn to remove precisely this friction. It may produce works that are increasingly consumable and fewer works that change the criteria of consumption. The danger is not obvious kitsch. It is quality without resistance.

Research on creativity already reveals a paradox. In experimental settings, access to generative ideas can improve the creative quality of individual outputs, especially for people who begin with fewer ideas, while reducing diversity across the collection. Each person can move away from his or her own baseline while many people move toward regions favored by the same generative distribution. Individual uplift and collective convergence can coexist (Doshi & Hauser, 2024).

This is the core of statistical culture. The term does not mean a world without originality. It describes a world in which the average quality floor rises dramatically while the distribution of recognizable forms becomes shaped by systems trained on previous distributions and commercial metrics. Models make the probable easy to produce. Recommendation systems make the

likely-to-engage easy to select. The combination can generate enormous surface variety while preserving deeper regularities.

Scientific culture is vulnerable to the same mechanism. Researchers can use common models to formulate questions, identify literature, propose methods, draft code, and write papers. Productivity rises, but collective inquiry can converge around the same visible problems and canonical sources. Recent commentary has warned about scientific monoculture around AI itself: rapid thematic convergence, shared methods, and institutional rewards for recognizable forms of novelty. The issue is not that models cannot suggest unusual ideas. It is that scientific institutions may reward the same kinds of legible surprise repeatedly (Traberg et al., 2026).

In science, diversity is not decoration. Competing methods and theoretical traditions protect against common error. A single model trained on the canon can help a newcomer enter a field while making omissions in the canon harder to see. If the same systems summarize evidence, generate hypotheses, and assist review, an epistemic blind spot can be reproduced at multiple stages with impressive consistency.

The answer is not to exclude AI from science or art. It is to design for divergence. Researchers and artists can deliberately request incompatible frames, use different models and corpora, separate exploration from polishing, preserve local archives, and compare machine-generated suggestions with sources outside dominant datasets. Institutions can reward unusual origins of

questions rather than only polished answers. Diversity requires process design once the cheapest path points toward convergence.

Maintaining diversity is costly. A dominant model, shared interface, and common metric improve interoperability and reduce training costs. Markets therefore have natural reasons to converge. Pluralism may require public funding, open models, local archives, independent institutions, and communities capable of refusing globally optimized criteria. Cultural diversity does not preserve itself merely because everyone says diversity is valuable.

Walter Benjamin described the loss of the artwork's aura under mechanical reproduction: the weakening of uniqueness tied to a particular time, place, and tradition. Generative AI extends the problem. It does not merely reproduce an object. It can reproduce traces of process, stylistic hesitation, texture, age, imperfection, and biography. Even apparent imperfection can be generated on demand (Benjamin, 1936).

For that reason, “made by a human” will not be sufficient as an aesthetic theory. A human-made work can be trivial, and a generated work can carry genuine meaning for a recipient. The more defensible concept of authenticity concerns the relation among a person, a process, and a claim. When someone says “this is my testimony,” the important fact is that the person stands behind the experience and can answer for it, not that no tool contributed to wording.

Nevertheless, human cost will become a signal. A live concert, in-person seminar, workshop, handwritten draft, bespoke service,

or object with documented provenance can acquire premium value because it demonstrates irreproducible time and presence. In an economy of instant generation, time becomes evidence. The cultural market may increasingly sell not only artifacts but proof that a person was there.

This renaissance of authenticity can be healthy and unequal. Wealthy people may purchase human teachers, doctors, artists, coaches, and communities, while everyone else receives competent synthetic services. “Human-only” can become a luxury analogous to artisanal production. Society will then face a difficult question: in which domains is human presence a premium preference, and in which is it part of equal citizenship?

Education, justice, care, and some forms of medicine cannot be treated simply as markets for authenticity. A poor child should not receive only automated teaching because it is inexpensive if reciprocal human mentorship has independent developmental value. A patient should not receive a human physician only as an elite upgrade. AI can expand access without building two regimes of humanity: relationship for those who can pay and simulation for those who cannot.

At the same time, generative systems can strengthen small cultures. A village, diaspora, minority language, research community, or local institution can produce translation, archives, educational material, and professional communication at a standard that previously required substantial capital. If the community controls data and governance, AI can increase

cultural sovereignty. If it merely applies local decoration over a globally optimized infrastructure, apparent diversity can conceal dependence.

Brand and community can become more important precisely because artifacts become cheap. In a market saturated with plausible content, audiences rely more heavily on trusted sources. Reputation reduces selection costs. This creates a paradox: decentralization of production can recentralize attention around a small number of names, platforms, institutions, and curators that provide trustworthy filters.

The opposite is possible as well. Communities can create shared brands, provenance systems, cooperative editorial institutions, and decentralized forms of reputation. Cheap production lowers the cost of constructing cultural infrastructure, not only content. The institutional question is whether those communities own the rules by which their production is validated and distributed.

Statistical culture therefore does not imply that everyone will look or sound identical. There may be more niches, symbols, genres, and identities than ever. Homogenization can occur at a deeper level: pacing, argumentative structure, notions of professionalism, interface expectations, criteria of attractiveness, and dependence on the same model families. Surface variety can coexist with infrastructural convergence.

This distinction matters politically. A system can produce perfectly customized messages for every social group while

optimizing the same commercial or behavioral objective. Personalization is not pluralism if every path leads to the same metric. A living culture possesses the ability to define different metrics and to preserve practices that are inefficient according to dominant platforms.

The role of the artist may therefore shift from producing every element toward constructing a world, choosing constraints, curating variants, and accepting responsibility for a final form. When variation is effectively infinite, stopping becomes a creative act. Curatorship is not necessarily less creative than manual execution, but it must mean more than choosing the statistically most effective option. Taste becomes the capacity to reject something good because something else is necessary.

Editorial work gains importance under the same conditions. The editor is not valuable because a machine cannot correct grammar. The editor protects structure, novelty, truth, proportion, and the relation between a work and an audience. Journalism becomes less valuable as summary and more valuable as access, verification, institutional memory, and reputational risk. Criticism competes poorly with cheap opinion generation but can become more valuable as the continuous record of an accountable judgment.

Provenance will become cultural infrastructure. Audiences will increasingly need to know what was generated, what was documented, which sources were used, who verified claims, and which version was encountered. Technical marking cannot

determine artistic value, but it can make informed choice possible. Without provenance, every object can claim any origin and authenticity becomes a purely promotional vocabulary.

An obsession with purity would be equally damaging. Reliable AI detection is limited, creative processes are already hybrid, and human culture has always depended on borrowing and tools. Dividing culture into "human true" and "AI fake" would flatten the very distinctions we need. The mature question is what relationship the object promises and whether that promise is honest. A collaboratively generated novel can be valuable. A fabricated testimony presented as lived experience creates a different ethical problem.

Cognitive inflation also changes the reader. In the near future, people will be able to consume more competent content than they can integrate. Better automatic filtering is not a complete solution because it strengthens the same optimization loop. Healthy response also requires deliberate limits: fewer trusted sources, communities with memory, reading rituals, editorial institutions, and times when selection is not delegated to an engagement model. Abundance creates a need for institutions of refusal.

The economic logic is important. When production is cheap, distribution and attention capture become more valuable. A creator can spend less on making an artifact and more on manipulating the channel through which it is encountered. This can worsen winner-take-most dynamics even while creation itself

democratizes. AI may create millions of competent producers and a smaller set of extraordinarily powerful distributors.

Culture can therefore move toward either smooth conformity or unprecedented plurality. Both futures are technically possible. The difference will be made by ownership, platform objectives, standards of provenance, public institutions, and the ability of communities to preserve forms that do not maximize immediate engagement. Originality will not disappear. It will have to survive the extraordinary efficiency with which recognizable novelty can now be manufactured.

When almost anything can be produced, the central cultural question changes. It is no longer merely "can we create this?" It becomes "why should this form occupy shared attention?" Generative AI is excellent at satisfying requests. Culture begins where individuals and communities decide which requests should not be left to the market, which difficulties are worth preserving, which common objects should remain common, and for what kinds of presence they are willing to spend irreplaceable human time.

CHAPTER VIII

# EDUCATION AFTER THE END OF PROOF

Modern education rests on a simple convention: the student's product is a trace of the student's mind. An essay indicates reading and organization of ideas. A solved problem indicates some understanding of method. A program indicates some algorithmic competence. The timed examination compresses these traces into a comparable measure. The convention has always been imperfect, but it was useful enough to distribute credentials and opportunities.

Generative AI does more than make cheating easier. It breaks the inference between artifact and learning even when use is fully permitted. One student can produce an excellent result, comply with every declared rule, and remain unable to explain the concepts. Another can use the same tool for feedback, comparison, and exploration and learn faster than before. The final file may not distinguish the two processes.

Adoption is already too broad to treat this as an edge case. The Stanford AI Index 2026 reports that roughly four out of five university students use generative AI tools. A general prohibition would turn normal practice into hidden practice and advantage students who are better at concealment. Uncritical acceptance would allow institutions to certify output rather than

competence. Education must give up the comfort of one universal policy (Stanford HAI, 2026).

Three objectives that schools often mix should be separated. The first is performance in the real world with tools. Here AI use is often realistic and necessary. The second is construction of internal competence that enables verification, transfer, and recovery. Here some unaided periods remain important. The third is certification: the institution must be able to make a credible public claim about what the person can do. Each objective requires a different form of evidence.

Assisted performance is not fake performance. Architects use software, programmers use libraries, researchers use databases, physicians use diagnostic instruments, and no serious theory of competence requires them to abandon these tools. In future professional life, excluding AI from every project would test an artificial world. Students must learn to formulate problems, compare outputs, verify sources, document uncertainty, integrate tools, and recognize when the system's contribution exceeds their ability to supervise it.

Yet competent use requires background knowledge that the tool can conceal. To recognize a wrong solution, a person needs models of the domain and experience with its variation. If every difficult stage is delegated before understanding develops, the student acquires a procedural ability to request outputs without the basis required for judgment. Prompting is not a universal

substitute for knowledge, just as access to search never made literacy unnecessary.

Early educational trials show the genuine potential of AI tutoring. In a randomized study, a carefully designed tutor produced stronger learning outcomes than an active classroom condition for specific material and did so in less time. The important detail is that the system did not function as an unrestricted answer generator. It operated inside a pedagogical structure with bounded content and feedback. The result does not establish that teachers are redundant. It shows that individualized explanation at scale can outperform some conventional arrangements on defined learning tasks (Kestin et al., 2025).

Teachers perform functions that immediate learning benchmarks do not capture. They observe development over time, create group norms, notice withdrawal and distress, model intellectual behavior, mediate conflict, and decide when difficulty is productive. A tutor can explain the same concept ten ways without always understanding why a student is avoiding the problem. The strongest architecture may therefore be a teacher coordinating artificial tutors and using recovered time for diagnosis, mentorship, and relationship.

Economic pressure may select a different architecture. Wealthy institutions can keep teachers and add AI as a multiplier. Poor institutions can replace part of the human relationship with standardized automated service. The second option may improve

test scores relative to no support at all and still create a stratification between education as human formation and education as individualized optimization. Society must decide which relational functions are public goods rather than luxury features.

Assessment has to be rebuilt around process, transfer, and reconstruction. Oral examination can require explanation, objection handling, and modification of a solution. An interactive exam can introduce new facts after the first answer. A practical task can be observed in stages. A portfolio can include decision logs, source checks, failed hypotheses, and later corrections. None is impossible to assist, but in combination they make absent competence harder to simulate.

These methods are more expensive than automatically grading an essay. That is precisely the political problem. When generation becomes cheap, credible certification becomes more expensive. Universities can answer with standardized surveillance-heavy examinations or with richer, smaller-scale evaluation. The first is scalable and can become coercive. The second is educationally richer and labor intensive. AI does not necessarily reduce the cost of education; it can move cost from information transmission toward observation of learning.

It is useful to distinguish assisted assessments from unaided resilience assessments. The assisted assessment measures realistic performance in the actual technological ecosystem. The resilience assessment asks whether the person can continue when the tool is

unavailable, misleading, or inappropriate. Aviation and medicine routinely train degraded modes and emergencies. Occasional unaided cognitive work can serve the same purpose without becoming a nostalgic ritual.

The residual competence required differs by field. Translators need not reproduce machine speed but must recognize shifts in meaning and register. Programmers need not memorize every API but must understand control flow, data, architecture, and security well enough to inspect generated systems. Physicians need not remember every publication but must detect when the recommendation does not fit the patient. Curriculum design should identify control competencies rather than preserve every historical task mechanically.

This is the basis of proof of cognition. The term does not mean inspecting private thoughts or demanding disclosure of hidden chains of reasoning. It means public evidence that a person can reconstruct relevant reasons, transfer a method, respond to variation, and defend a decision. Students need not prove that no machine helped them. They need to prove that they are more than carriers of an artifact.

Proof of cognition must itself be bounded. Institutions can turn every learning activity into an examination and every interaction into a permanent record. Students need spaces in which they can experiment, make mistakes, use tools, and ask naive questions without each trace becoming part of an evaluative dossier. The separation between practice and certification is as

important in education as the separation between assistance and employee evaluation at work.

AI can support assessment by generating adaptive questions, counterexamples, simulations, and feedback. But the system that helped the student should not be the only judge of competence. Automated evaluation can privilege styles, dialects, argument forms, and assumptions represented in training data. Apparent objectivity can standardize academic culture and punish difficult originality.

The teacher's role therefore shifts away from judging polish alone and toward judging structure. Generated prose can be stylistically impeccable and conceptually empty. Awkward prose can contain an original insight. Historically, form carried some information about effort; now it can mask its absence. Human pedagogical judgment becomes more important precisely when administrations may be tempted to automate it.

Writing should remain part of education even if machines write better, because formulation is a cognitive process. But "writing without AI" should not become an absolute value. Some stages can be assisted without losing the core: spelling, language support, formatting, preliminary organization. At other stages, immediate generation closes the question before the student has encountered it. Pedagogy should therefore care about sequence: first hypothesis, then assistant; first outline, then critique; first explanation, then comparison.

Programming makes the dilemma concrete. Coding agents can produce functioning applications quickly, giving students the satisfaction of an outcome without requiring understanding of state, concurrency, data flow, dependencies, or security. Total prohibition prepares students poorly for contemporary practice. Total delegation can turn them into operators of systems they cannot debug. A serious curriculum will mix agentic projects with code review, explanation, architecture reconstruction, adversarial debugging, and exercises in which generated code contains subtle defects.

The humanities face a different risk. An essay is not merely a container for a conclusion; it is a discipline of inhabiting an argument long enough to discover its tensions. If the model immediately supplies the most probable structure, the student can avoid the encounter with ambiguity that the assignment was designed to create. Yet AI can also stage objections, compare interpretations, translate difficult passages, and make obscure contexts accessible. Used as adversary and interlocutor rather than ghostwriter, it can intensify reading.

Mathematics and science reveal another distinction. A system can produce a correct answer while concealing the dependency structure that makes the result valid. Students therefore need experience tracing assumptions, units, error propagation, and proof obligations. In science, generated hypotheses are educationally useful only if students learn how evidence can

falsify them. The cheapness of proposing explanations increases the importance of teaching how explanations die.

Children require the greatest caution because foundational capacities are still developing. Learning is not only information transfer; it is development of attention, language, self-regulation, social trust, and tolerance for frustration. An infinitely patient tutor can be transformative for a child with little support and can also become an avoidance device for real relationships. Products for children should not be optimized primarily for engagement or commercial attachment.

At the same time, absolute parental surveillance is not a good solution. Children and adolescents need some private space for exploration. A safe architecture must protect them from commercial exploitation and severe risk while allowing development of an interior life that does not become a permanent dossier. Memory controls, age-appropriate transparency, limited anthropomorphism, and pathways to responsible adults will matter as much as content filters.

Adolescents may benefit especially from a non-judgmental space for questions about identity, sexuality, fear, failure, and social conflict. The benefit is real because embarrassment blocks help-seeking. Yet an artificial confidant can also become an unusually influential participant in identity formation. The appropriate standard is not that the system never discusses sensitive matters, but that it is grounded, privacy-preserving, clear

about its nature, and connected to credible human support when needed.

The educational institution must also reconsider what credentials mean. A degree can no longer plausibly certify that every polished artifact was produced unaided. It can instead certify a layered capability: foundational knowledge, unaided resilience, ability to work with advanced tools, verification competence, and ethical understanding of delegation. Such credentials may be more informative than the old fiction of pure individual production.

This change could restore some older forms of education that mass systems abandoned because they were expensive. Oral defense, apprenticeship, seminar discussion, supervised practice, and public demonstration become more valuable when take-home artifacts are cheap. AI may paradoxically make education more human at the point of certification, because machine generation forces institutions to observe the learner more directly.

Or it may produce the opposite: continuous automated surveillance intended to prove that every keystroke was authentic. The second path is technically easy and educationally dangerous. A student who is constantly monitored learns to perform compliance rather than think. The goal should be robust evidence at selected moments, not total visibility of intellectual life.

The economics of higher education will be disrupted accordingly. If explanation and basic tutoring become cheap,

universities must justify the value of laboratories, peers, mentorship, trusted evaluation, research communities, and social networks. Institutions that were largely delivering lectures may face strong pressure. Institutions that create demanding communities of practice may become more valuable even if much of their content is available freely.

The division between elite and mass education is therefore one of the central political risks of AI. Personalized instruction could be the greatest democratization of education since mass literacy. It could also become a reason to reduce human contact for those with the least power. A defensible policy should treat AI as a floor-raising infrastructure while protecting access to human mentorship where human relationship is part of the educational outcome.

Education after the end of the artifact-as-proof will be more difficult to administer because it must state explicitly what was once inferred cheaply. What does the learner know? What can the learner do with tools? What can the learner do without them? Can the learner tell when the tool is wrong? Can the learner explain a decision to another person? Can the learner transfer knowledge into a new situation? These questions are more demanding than asking for an essay, but they are closer to what education has always claimed to certify.

AI does not make education obsolete. It exposes a weakness that was already there: institutions often confused the production of school artifacts with the development of minds. Once artifacts

can be manufactured cheaply, the confusion can no longer be hidden. The opportunity is to redesign education around actual capability. The danger is to replace one proxy with another and call the resulting efficiency learning.

CHAPTER IX

# INTIMATE PROPAGANDA AND THE CONVERSATIONAL STATE

Political communication has always depended on the available architecture of media. The pamphlet enabled rapid polemic, the newspaper created recurring publics, radio synchronized national audiences, television transformed visual familiarity into political capital, and social platforms made ranking and micro-targeting central. Generative AI changes the next variable: it lowers the cost of producing an adaptive persuasive message during the interaction itself.

The difference between broadcast and conversation is politically important. Radio could transmit one speech to ten million listeners. Digital advertising could select different prepared messages for different groups. A generative system can, in principle, conduct ten million different conversations, learn from each objection, and alter framing in real time. The political object is no longer only a message. It is a responsive process.

Experimental evidence suggests that this capacity deserves serious attention. Pre-registered studies have found that language models can produce political persuasion comparable to, and under some conditions stronger than, human interlocutors. Personalization using basic information about the recipient can

increase persuasive effectiveness. These effects are not mind control, and they vary by topic and condition. They demonstrate something more modest but historically significant: individualized interactive persuasion can be produced at very low marginal cost (Salvi et al., 2025; Bai et al., 2025).

This is intimate propaganda. Classical propaganda sought slogans broad enough to mobilize masses. Intimate propaganda can use the language, examples, fears, moral foundations, and identities of one person. It can avoid arguments that would provoke resistance and emphasize truths likely to matter to that recipient. The danger therefore cannot be reduced to misinformation. A system could use only factually correct statements while manipulating selection, omission, sequence, and emotional framing.

The concept also changes the meaning of public discourse. Democratic argument traditionally assumes, at least imperfectly, that reasons can be exposed to other citizens. A politician gives a speech; journalists quote it; opponents respond. Micro-targeted advertising already weakens this common visibility. Generative conversation can fragment it further. Two voters can receive different justifications for the same policy, each optimized to be acceptable to that individual, with no stable public object against which inconsistency can be challenged.

This does not mean political personalization is inherently illegitimate. Citizens differ in language, education, disability, and prior knowledge. A tax reform should be explainable differently to

a teenager, a small-business owner, and an economist. The democratic requirement is not identical wording. It is consistency of the underlying public reasons and the ability to audit whether personalization changed explanation or changed the substance of the justification.

The same architecture can therefore support a conversational state. Public administration is difficult partly because rules are written for legal precision rather than ordinary comprehension. A citizen-facing agent can explain eligibility, deadlines, documents, rights, and appeal procedures in accessible language. It can translate, adapt to disability, and guide people through complex services. For citizens who cannot afford lawyers or consultants, this can be an extraordinary increase in practical access to the state.

The administrative benefit should not be underestimated. Bureaucratic complexity systematically advantages people with education, time, money, and institutional familiarity. A high-quality public agent could reduce that inequality. It could also expose contradictions in procedures, summarize the status of a case, and make public services more legible. In this sense, AI can act as an interface of administrative justice.

Yet the conversational state contains the same asymmetry as the commercial companion. The system may know far more about the citizen than the citizen knows about the system. It can remember previous cases, infer vulnerabilities, and select explanations. If the same infrastructure decides and explains, the

explanation can become a persuasive defense of its own decision rather than an independent account. A friendly interface can conceal a rigid or unfair rule.

This is why public reasons must remain public. Personalized explanation may vary in vocabulary and detail, but consequential decisions should be traceable to rules, evidence, and procedures that other citizens, journalists, courts, and auditors can inspect. The right to an explanation is insufficient if the explanation is merely another generated text. What matters is the ability to connect that text to an authoritative decision path and challenge it.

Appeal must also be genuinely independent. A button labeled "request review" that sends the same data back to the same model is not meaningful contestation. High-stakes systems need institutional routes through which a different decision maker, model, or human body can reassess the case. The architecture of appeal is part of democratic legitimacy.

Generative systems also lower the cost of political participation. Citizens can summarize legislation, draft petitions, analyze budgets, compare manifestos, prepare questions for representatives, and translate technical documents. Civil-society organizations with limited resources can acquire research and communication capabilities that once belonged to governments, parties, and large NGOs. This can widen participation and strengthen oversight.

But cheap participation produces its own inflation. If every citizen can generate a thousand formal complaints, consultation submissions, or messages to legislators, public institutions can be overwhelmed. Representatives may respond with automated summarization and generated replies. Political communication becomes machine-to-machine at the edges: agents write to agents, summarize agents, and produce a thin layer of human attention at the end. Volume no longer proves intensity of preference.

Democratic systems will therefore need stronger signals of commitment. Identity, deliberative participation, time spent, verified constituency, randomly selected citizen assemblies, or structured argument may become more important than raw message counts. Once expression is nearly free, institutions need ways to distinguish mass-generated noise from durable public concern without excluding people who need automated assistance to express themselves.

Elections face a related problem. Synthetic media make impersonation, fabricated events, and mass production of campaign material cheaper. Provenance standards, rapid correction, trusted channels, and media literacy become more important. Yet deepfakes may not be the most consequential effect. Personalized persuasion that is technically truthful, legally compliant, and invisible to outsiders can be harder to govern because it does not produce an obvious fake object to debunk.

Political parties can use AI to simulate voter segments, draft policy explanations, canvass through conversational agents, and

optimize outreach. Such tools can make small parties more competitive and can also increase the power of organizations holding superior data. The key scarce resource becomes not the ability to write a persuasive message but access to detailed behavioral profiles and permission to use them.

This creates a convergence between privacy and democracy. Personal data are not merely commercial assets when they allow political systems to discover which argument is most effective for which person. A society can protect ballot secrecy while permitting unprecedented pre-vote psychological targeting. Data governance therefore becomes part of the architecture of political autonomy.

There is also a geopolitical layer. States can use generative systems for influence operations at a scale and linguistic diversity previously requiring large human teams. Automated personas can participate in local debates, translate cultural references, and maintain long-term interaction. At the same time, detection and counter-disinformation systems can also become stronger. The result is an arms race in synthetic credibility rather than a one-directional victory of attackers.

The history of radio is useful here. Propaganda effects depended on institutional control and prior local beliefs. Generative persuasion will also interact with trust, identity, media systems, and political culture. A population with credible institutions and plural information sources may be more resilient than a population in which all sides already assume that every

institution lies. AI can exploit distrust, but distrust usually has a history.

The political danger of AI is therefore not that it creates ideology from nothing. It can reduce the cost of finding and amplifying fractures that already exist. It can test narratives rapidly and adapt them across communities. Polarization may emerge not from one common propaganda message but from many locally optimized messages that make compromise appear irrational to each group for different reasons.

At the same time, AI could improve deliberation. Systems can steelman opposing arguments, identify factual disagreements, reveal where values rather than facts differ, and translate between vocabularies used by hostile groups. A deliberative assistant optimized for mutual understanding would be technologically close to a persuasive assistant optimized for victory. The distinction lies in objective, governance, and transparency.

This makes alignment political rather than merely technical. A model that is “helpful” must be helpful toward some conception of the user's objective. A public model may have to balance user preference with legality, factuality, non-discrimination, institutional duty, and the interests of people affected by the user's actions. There is no purely neutral optimization once the system participates in social decisions.

Governments themselves will increasingly use AI internally. OECD data already show broad adoption for administrative and internal processes, with more caution around policymaking and

accountability functions. That difference is rational. Summarizing case files and scheduling inspections are different from defining rights or choosing policy trade-offs. The boundary will nevertheless move as systems become more capable (OECD, 2025/2026).

The European Union's AI regulatory framework emphasizes transparency, risk management, and obligations around certain AI interactions and generated content. The United States has, in recent policy, placed comparatively greater emphasis on innovation, infrastructure, and geopolitical leadership. Other states will construct still different balances among control, industrial policy, surveillance, and public service. There will not be one global political settlement around AI.

These differences matter because AI infrastructure has scale economies. If a few companies or states control frontier models, compute, identity layers, and personal data, political autonomy can become dependent on private infrastructure. Public institutions may need procurement rules, open standards, portability, independent evaluation, and in some cases public or sovereign capacity not because every state should build everything, but because legitimate government cannot become technologically incapable of changing suppliers.

A conversational state also needs a constitutional rule about memory. Citizens should not have to repeat their situation endlessly to every agency, but a perfectly integrated state memory can become an infrastructure of surveillance. Convenience and

civil liberty point in opposite directions. Data minimization, purpose limitation, compartmentalization, expiration, and user-visible access logs can become constitutional principles of the interface.

The ideal system would be asymmetrically transparent: the citizen can understand what the state knows and why it acts, while the state collects only what it legitimately needs. Current digital systems often approximate the opposite. AI raises the stakes because inferred information can be as consequential as explicitly provided information.

Another political issue is legitimacy of automated expertise. Governments already rely on models, forecasts, and expert agencies. AI can widen expert capacity but also make technocratic language more persuasive. A generated policy memo can present trade-offs with extraordinary coherence, creating the impression that one option follows naturally from evidence. Yet many political decisions concern distribution among legitimate interests rather than factual optimization. The system can clarify the choice; it cannot convert the choice into a neutral calculation.

Democratic institutions therefore need to preserve the distinction between prediction and authorization. A model can estimate which policy will maximize a metric. Citizens and representatives must still decide whether the metric should dominate. Treating predicted effectiveness as legitimacy would repeat an old technocratic error in a more fluent form.

The optimistic political future is not one in which AI governs. It is one in which citizens gain inexpensive expertise, public institutions become understandable, participation becomes easier, and decision paths become more auditable. The pessimistic future is not necessarily a dictatorship run by machines. It can be a formally democratic society in which persuasion is individualized, public reasons fragment, administrative systems become impossible to contest, and essential cognitive infrastructure is owned elsewhere.

The difference will be determined by architecture before it is determined by rhetoric. Who controls memory? Who can personalize? Which objectives are legal to optimize? Can citizens inspect the rule behind the answer? Are alternative providers possible? Does appeal reach an independent authority? Are public explanations archived in a form that allows comparison? These design questions are democratic institutions in technical form.

The great political challenge of generative AI is therefore to preserve a common world while permitting individualized assistance. Democracy requires citizens who can receive explanations appropriate to them, but it also requires reasons that remain shareable enough to be contested together. The conversational state will be legitimate only if personalization improves access without privatizing the grounds of public power.

CHAPTER X

# THE BODY, MEDICINE, AND THE RIGHT TO REMAIN CAPABLE

The body is the place where exaggerated claims about AI become hardest to sustain. A benchmark can improve dramatically while a patient remains sick. A diagnostic model can outperform a clinician on a defined task while a hospital fails because the drug is unavailable, the patient cannot return, the record is incomplete, or the organization routes the case badly. Medicine forces a distinction that the rest of society often avoids: performance of a cognitive component is not performance of the entire system.

This distinction was visible in a pragmatic randomized trial published in Nature Medicine in 2026. Clinicians in Kenyan primary care using an AI-supported system improved documentation and aspects of recorded diagnosis and treatment planning, yet the primary patient outcome, treatment failure at fourteen days, did not improve significantly. The result is neither a failure of AI nor a triviality. It shows why real-world evaluation must follow the causal chain all the way to what society actually cares about (Agweyu et al., 2026).

The lesson generalizes beyond medicine. Better legal drafting does not automatically produce justice. Better educational output

does not automatically produce learning. Better software generation does not automatically produce reliable systems. Better administrative summaries do not automatically produce legitimate decisions. AI is often evaluated at the level where it is easiest to measure, while the social objective exists several institutional layers downstream.

Medicine also exposes the irony of automation because the human body produces rare and consequential exceptions. Clinical systems can automate ordinary pattern recognition, documentation, triage, and monitoring. Humans then encounter a larger proportion of the ambiguous cases. The better the system performs on routine situations, the less clinicians practice some of the skills for which they are retained. High reliability can therefore create low readiness for unusual failure.

A multicenter study of colonoscopy has intensified concern about this mechanism by reporting lower unassisted detection performance after clinicians had been exposed to AI-assisted detection. The result must be interpreted carefully and should not be generalized to every clinical tool. But the mechanism is familiar from aviation and industrial automation: the system takes over the signal continuously, human practice declines, and the human is asked to recover precisely when the system is absent or wrong (Budzyń et al., 2025; Bainbridge, 1983).

This is why medical automation cannot be governed only by average accuracy. We need measures of residual capability, failure recognition, degraded-mode performance, distribution of errors,

and effects on training over time. A tool can be beneficial to today's patients and still damage tomorrow's workforce if it consumes expertise without reproducing it. Conversely, a system can improve training by exposing clinicians to rare cases, giving feedback, and making reasoning visible. The difference is institutional design.

Clinical responsibility provides another test. If a physician must legally answer for a recommendation generated by a system, the physician needs enough information and authority to reject it. Liability without epistemic control is unfair and unsafe. But transferring all responsibility to the manufacturer is also incomplete because clinical decisions depend on local context. The correct allocation is likely to be distributed but explicit: responsibilities attached to model design, institutional deployment, professional use, and patient-specific judgment.

This logic returns us to the constitutional organization. In high-stakes medicine, intention, generation, verification, and authorization should be separable. A model can propose a diagnosis; a different process can check contraindications; the clinician can integrate patient preference; institutional protocols can control exceptional risk. The goal is not maximal human involvement in every micro-step. It is enough independent structure that no single failure mode silently becomes the decision.

AI may also reshape the doctor-patient relationship in positive ways. Documentation is a major source of administrative burden.

Systems that reduce note-taking can return visual attention and conversation to the patient. Translation can improve access. Summaries can help people understand complex histories. Rare-disease information can reach primary care. A clinician supported by systems may have more time to exercise specifically human functions that were crowded out by paperwork.

The opposite outcome is equally plausible. Organizations can convert every minute saved into higher throughput. The physician sees more patients, the system produces more notes, and the encounter becomes even more compressed. Technology creates a time dividend; payment models and management decide whether it becomes care or acceleration. The medical example shows why productivity gains have no inherent social distribution.

Mental health makes the ambiguity still sharper. Conversational systems can provide immediate support at a scale that human services cannot currently match. A 2026 meta-analysis of randomized trials found small but statistically significant reductions in depressive and anxiety symptoms across chatbot interventions, with larger effects for depression in some clinical and subclinical populations. At the same time, heterogeneity was high, many studies carried substantial risk of bias, and technologies and protocols varied greatly. The evidence supports careful use as support or bridge, not automatic equivalence with psychotherapy and not dismissal as meaningless simulation (Sohn et al., 2026).

Access changes the ethical baseline. For a person waiting months for professional help, an imperfect intervention can be valuable. For a health system tempted to replace adequate human care with cheaper automation, the same intervention can become a mechanism of inferior provision. There is no context-free category called "good therapeutic AI." Value depends on the alternative, severity, privacy, escalation pathways, clinical integration, and business model.

These examples also discipline claims about biology. There is currently no robust basis for saying that ordinary LLM use causes generalized neurological deterioration. Public discussion often jumps from studies showing different patterns of neural activity or attention during technology use to claims of brain damage. Such inference is not justified. Neuroplasticity means that repeated activity changes patterns of processing and skill. It does not mean that every change is pathological.

Research on short-form video is relevant because it demonstrates that media architecture can affect cognition beyond content. Studies using behavioral, eye-tracking, and neurocognitive measures have associated highly fragmented short-video exposure with changes in memory, attention, integration, and control under particular experimental conditions. The results remain context-dependent and do not describe the destiny of all users. They show that the temporal organization of a medium is part of what the nervous system experiences (Li et al., 2025; Wei et al., 2026).

For AI, the most plausible biological effects in the near future are indirect. Which cognitive activities do people stop practicing? How much time moves from human conversation to artificial interaction? How do sleep, posture, movement, stress, and work tempo change? An assistant that removes administrative burden can reduce stress. A system that makes twenty-four-hour productivity possible can increase it. The same technology can protect one worker's body and accelerate another's toward exhaustion.

The medicalization of normal adaptation should also be resisted. If people memorize fewer facts because external retrieval is reliable, this is not automatically cognitive disease. If spelling, translation, or arithmetic become less practiced, society must ask whether the lost capacity was important for autonomy, development, or recovery. Biological preservation is not an end in itself. The relevant question is whether the human-system arrangement remains healthy, resilient, and under meaningful control.

Children are the largest unknown because no longitudinal evidence can yet cover a generation raised from early childhood with permanent generative interlocutors. We can study early tutoring, affective conversation, and willingness to disclose, but we do not know how attention develops when explanations are instant, how frustration tolerance changes when the partner adapts endlessly, or how reciprocity is learned when one stable conversational relationship may be a commercial service.

The absence of evidence does not justify panic, and it does not justify laissez-faire. Children are not small adults with different privacy settings. Systems designed for them should have objectives other than maximizing engagement, controllable memory, clear limits on anthropomorphic deception, age-appropriate boundaries, and pathways to responsible adults. Yet total parental visibility can also destroy the private space through which identity develops. Protection from exploitation must coexist with a right not to have every thought permanently recorded.

Biological evolution in the genetic sense is probably less important over the next several decades than cultural evolution. Human genomes change slowly compared with product cycles. By 2050 it is more plausible that we will see different cultural cognitive phenotypes than a biologically transformed species: people who remember relatively little but coordinate external systems superbly; people who maintain unaided capacities deliberately; people whose lives are almost completely mediated; and people who purchase periods of technological absence as a luxury.

The body nevertheless places a limit on narratives of frictionless cognition. Attention is finite. Sleep is required. Emotional regulation has physiological costs. Human social interaction depends on timing, gesture, touch, and co-presence that text models only partially represent. AI can expand symbolic capacity by orders of magnitude without multiplying biological time. This mismatch may become one of the defining constraints

of the century: almost infinite possible output competing for a nervous system that still inhabits twenty-four-hour days.

Cognitive inflation can therefore become a medical problem indirectly. More messages, more tasks, more generated opportunities, and more personalized stimulation can increase demands on attention even when each task is individually easier. A worker may perform less cognitive labor per artifact and process far more artifacts. The experience of overload can grow inside an economy celebrating automation.

The same applies to social comparison. If synthetic systems produce polished language, bodies, images, careers, and personas at enormous scale, standards of normal performance can shift. Social media already made comparison with selected lives routine. Generative media can create comparison with lives that never existed. Effects on anxiety, body image, and self-evaluation will depend heavily on design and context, but the mechanism deserves monitoring rather than retrospective surprise.

At this point the threads of the book converge. Medicine shows that a good symbolic result does not guarantee a good bodily result. Deskilling shows that assisted performance does not guarantee residual capacity. Psychology shows that production does not guarantee authorship. Education shows that an artifact does not guarantee learning. Politics shows that persuasion does not guarantee a public reason. Across domains, the old inference from outcome to competence, understanding, or legitimacy becomes less reliable.

The normative response can be summarized as the right to remain capable. This is not a right to reject every technology, nor an obligation to preserve every skill of the past. It is the claim that persons and institutions should retain enough independent capacity to understand what is at stake, detect important failure, challenge consequential decisions, change providers or procedures, and continue under reasonable degraded conditions.

At the individual level, the right means access to education that does not confuse tool use with learning. It includes opportunities to practice selected abilities without assistance, access to underlying sources, portability of personal data and memory, and the ability to change systems without losing a large part of one's accumulated cognitive identity. Good design should leave users more capable over time, not merely produce better outputs in each session.

At the professional level, the right implies preserved apprenticeship, failure simulation, and measurement of unaided competence where failure is consequential. An organization that automates junior work should fund alternative pathways through which juniors encounter variation, make supervised mistakes, and build mental models. Otherwise it consumes the expertise accumulated by the current generation and hands the next generation only an interface.

At the organizational level, remaining capable means contestability. Employees, clients, patients, and citizens should be able to know when an important decision was materially

delegated, what evidence shaped it, and who possesses genuine authority to change it. There must be a person or institution with real power, not an appeal form that routes the case back through the same mechanism. Responsibility cannot be distributed until it disappears.

At the societal level, the principle requires plural suppliers, open standards, independent evaluation, public-interest infrastructure, and communities of expertise capable of auditing dominant systems. No society is cognitively autonomous if every critical institution depends on a small number of systems that it cannot replace or meaningfully inspect. Cognitive sovereignty does not require technological autarky. It requires credible options and bargaining capacity.

At the psychological and biological level, remaining capable includes protection of attention and development. Not every environment should be optimized for retention. Sustained attention, reciprocal conversation, movement, sleep, privacy, and periods without adaptive stimulation can be understood as health infrastructure. This does not justify moral panic about entertainment. It recognizes that markets can exploit reward mechanisms more quickly than individuals can develop defenses.

The principle has an important limit. Capability must not become a new cruel meritocracy. People who use cognitive prostheses, translation, disability assistance, or medical support are not less authentic. Autonomy has never meant absolute independence. Human beings depend on language, other people,

institutions, and technology. What should be protected is negotiable dependence: the ability to understand the relation, change suppliers, request human assistance, and participate in the rules that govern the dependency.

Several futures remain plausible and can coexist. In one, AI becomes infrastructure for cognitive abundance: mid-level expertise spreads, services become accessible, and people use saved time for more difficult problems. In another, cognitive bifurcation deepens: a minority amplifies judgment while a majority receives excellent results without the capacity to verify them. In a third, institutions become extraordinarily productive and fragile, filled with artifacts no one fully understands. In a fourth, cultural reaction creates protected spaces for unaided education, human relationship, local institutions, and authenticated presence.

None of these futures is contained inside the model. Technical capability opens a space of possibilities; incentives push through that space; institutions give the trajectory form. Printing did not contain the Reformation, radio did not contain fascism, television did not contain one inevitable mass culture, and the early internet did not mechanically contain today's platform economy. Each technology changed costs and possibilities, after which societies built concrete regimes, often only recognizing the political stakes after dominant interests had become difficult to challenge.

AI lowers the cost of plausible symbolic cognition. That is both promise and danger. It can place a tutor beside every child, a

consultant beside every clinician, and an advocate beside every citizen. It can also place a persuasive agent beside every vulnerability, a generator beside every metric, and a layer of fluent explanation over institutions that no longer understand what they are doing.

The Great Decoupling should not be stopped wholesale. Many old couplings were unjust. Effort was not rewarded fairly. Expertise was scarce and expensive. Disability, language, geography, and class excluded capable people. AI can break real barriers. But decoupling must be followed by deliberate recoupling: action with responsibility, assistance with learning, power with contestability, efficiency with resilience, and cognitive abundance with preservation of the capacities required to govern it.

The central question is not whether humans will remain superior to machines at every task. They almost certainly will not, and the comparison is too narrow. The question is whether the hybrid systems we build will keep human beings as capable participants or reduce them to sources of approximate objectives, behavioral data, and legal responsibility. A civilization can become far more intelligent than any of its members without making those members powerless inside it.

The right to remain capable is, finally, the right not to become the competent spectator of a world that acts in one's name. It does not demand that we do everything alone. It demands that we can understand enough, say no, learn again, choose another route,

and find other people with whom the system can be reconstructed when fluent answers cease to correspond to reality. That is the threshold between successful externalization of cognition and surrender to it.

# SELECTED BIBLIOGRAPHY


Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of the Nazis in Prewar Germany. The Quarterly Journal of Economics, 130(4), 1885-1939. doi:10.1093/qje/qjv030.

Agweyu, A., et al. (2026). Generative AI-enabled clinical decision support system in primary care: A pragmatic, cluster-randomized trial. Nature Medicine. doi:10.1038/s41591-026-04503-6.

Andersson, M. (2025). Companionship in code: AI's role in the future of human connection. Humanities and Social Sciences Communications, 12, 1177. doi:10.1057/s41599-025-05536-x.

Bai, H., et al. (2025). LLM-generated messages can persuade humans on policy issues. Nature Communications, 16, 6037. doi:10.1038/s41467-025-61345-5.

Bainbridge, L. (1983). Ironies of Automation. Automatica, 19(6), 775-779. doi:10.1016/0005-1098(83)90046-8.

Benjamin, W. (1936/2008). The Work of Art in the Age of Its Technological Reproducibility. In M. W. Jennings, B. Doherty, & T. Y. Levin (Eds.), The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media. Harvard University Press.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. The Quarterly Journal of Economics, 140(2), 889-942. doi:10.1093/qje/qjae044.

Budzyń, K., et al. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. The Lancet Gastroenterology & Hepatology, 10(10), 896-903. doi:10.1016/S2468-1253(25)00133-5.

Clark, A. (2025). Extending Minds with Generative AI. Nature Communications, 16, 4627. doi:10.1038/s41467-025-59906-9.

Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58(1), 7-19. doi:10.1093/analys/58.1.7.

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects

of Artificial Intelligence on Knowledge Worker Productivity and Quality. Organization Science, 37(2), 403-423. doi:10.1287/orsc.2025.21838.

Dittmar, J. E. (2011). Information Technology and Economic Change: The Impact of the Printing Press. The Quarterly Journal of Economics, 126(3), 1133-1172. doi:10.1093/qje/qjr035.

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. doi:10.1126/sciadv.adn5290.

Durante, R., Pinotti, P., & Tesei, A. (2019). The Political Legacy of Entertainment TV. American Economic Review, 109(7), 2497-2530. doi:10.1257/aer.20150958.

European Commission. (2026). Transparency obligations under Article 50 of the AI Act: Questions and answers. Directorate-General for Communications Networks, Content and Technology.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. Official Journal of the European Union.

Glickman, M., & Sharot, T. (2025). How human-AI feedback loops alter human perceptual, emotional and social judgements. Nature Human Behaviour, 9, 345-359. doi:10.1038/s41562-024-02077-2.

Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). Generative AI and Jobs: A Refined Global Index of Occupational Exposure. ILO Working Paper 140. International Labour Organization. doi:10.54394/HETP0387.

Guess, A. M., et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? Science, 381(6656), 398-404. doi:10.1126/science.abp9364.

Humlum, A., & Vestergaard, E. (2025). Early Labor Market Transformation under Generative AI (previously circulated as Large Language Models, Small Labor Market Effects). NBER Working Paper 33777. National Bureau of Economic Research. doi:10.3386/w33777.

Kestin, G., Miller, K., Klales, A., et al. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an

authentic educational setting. Scientific Reports, 15, 17458. doi:10.1038/s41598-025-97652-6.

La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. American Economic Journal: Applied Economics, 4(4), 1-31. doi:10.1257/app.4.4.1.

Lee, E. H., Yin, Y., Jia, N., & Wakslak, C. J. (2026). Relying on AI at work reduces self-efficacy, ownership, and meaning while active collaboration mitigates the effects. Scientific Reports, 16, 13583. doi:10.1038/s41598-026-42312-6.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Article 1121, 1-22. doi:10.1145/3706598.3713778.

Li, H., Li, J., Hao, X., & Liu, W. (2025). Behavioral and eye-tracking investigation of event segmentation following short video watching. npj Science of Learning, 10, 86. doi:10.1038/s41539-025-00378-3.

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nature Human Behaviour, 7, 74-101. doi:10.1038/s41562-022-01460-1.

Meincke, L., Nave, G., & Terwiesch, C. (2025). ChatGPT decreases idea diversity in brainstorming. Nature Human Behaviour, 9(6), 1107-1109. doi:10.1038/s41562-025-02173-x.

OECD. (2025). Generative AI and the SME Workforce: New Survey Evidence. OECD Publishing. doi:10.1787/2d08b99d-en.

OECD. (2026a). Digital Government Outlook 2026: From Foundations to Transformational Impact. OECD Publishing. doi:10.1787/0496b2bc-en.

OECD. (2026b). Generative AI Experimentation in Government: Learning from Emerging Guidelines. OECD Working Papers on Public Governance, No. 93. OECD Publishing. doi:10.1787/42815683-en.

OpenAI & MIT Media Lab. (2025). Investigating Affective Use and Emotional Well-being on ChatGPT. Research report and pre-registered study.

Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2025). On the conversational persuasiveness of GPT-4. Nature Human Behaviour, 9, 1645-1653. doi:10.1038/s41562-025-02194-6.

Sohn, J.-S., Ha, B.-G., Park, S., et al. (2026). Systematic review and meta analysis of chatbots in the management of depressive and anxiety symptoms. npj Digital Medicine, 9, 377. doi:10.1038/s41746-026-02566-w.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science, 333(6043), 776-778. doi:10.1126/science.1207745.

Stanford HAI. (2026). AI Index Report 2026. Stanford Institute for Human-Centered Artificial Intelligence, Stanford University.

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2026). AI is turning research into a scientific monoculture. Communications Psychology, 4, 37. doi:10.1038/s44271-026-00428-5.

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8(12), 2293-2303. doi:10.1038/s41562-024-02024-1.

Wei, M., Liu, J., Wang, H., Li, Q., et al. (2026). Fragmented learning from short videos modulates neural activity and connectivity during memory retrieval. npj Science of Learning, 11, 15. doi:10.1038/s41539-025-00399-y.

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X.-Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Scientific Reports, 15, 15105. doi:10.1038/s41598-025-98385-2.

Zhou, J., Corbett, F., Byun, J., Porat, T., et al. (2026). A systematic review and meta-analysis of psychological and behavioural responses in human-agent vs. human-human interactions. Communications Psychology, 4, 102. doi:10.1038/s44271-026-00466-z.