POLICY PAPER • 2026

HUMAN AGENCY IN THE AGE OF AI

A Strategic Framework for Cognitive Resilience and Human Flourishing

By Stephen C. Jordan

Executive Director, Institute for Sustainable Development | Alexandria, Virginia

A NOTE TO INDUSTRY READERS: This paper argues that AI should be deployed in ways that increase human capability, strengthen institutional resilience, and preserve human agency over consequential decisions. The AI industry has a strong commercial interest in this outcome: public trust is a prerequisite for broad adoption, and trust is harder to earn than capability is to demonstrate. Andrew Carnegie invested his industrial fortune in the libraries, universities, and civic institutions that equipped a new generation for a new economy. The frameworks proposed here — for lifelong learning, workforce transition, and community investment — are not regulatory burdens. They are the modern version of Carnegie's program, and the condition for sustained public legitimacy.

EXECUTIVE SUMMARY

Artificial intelligence is already improving scientific work, operational efficiency, and decision support across the economy. It also creates governance obligations of historic scale. The central question is not whether AI should be adopted, but how — in ways that expand human capability rather than displace human judgment, distribute benefits broadly rather than concentrate them narrowly, and equip society for a transition more disruptive and more prolonged than any previous technological transformation in American history.

This paper introduces the concept of intellectual protectionism: a policy posture that preserves human cognitive capacity and agency in domains where AI failure could be catastrophic, irreversible, or constitutionally significant. But governance frameworks alone are insufficient. The AI transition requires a social compact — a coordinated set of affirmative investments by industry, government, and civil society — without which political backlash will undermine the frameworks themselves.

The framework rests on four arguments. The hollowing-out analogy shows that America's offshoring of manufacturing destroyed not just jobs but the ecosystems of tacit knowledge that made industries governable — AI adoption risks the same at the level of human cognition. The automation complacency evidence demonstrates that skill atrophy from over-reliance on automated systems is empirically documented and potentially catastrophic in high-stakes domains. The trust problem establishes that voluntary industry self-governance is structurally insufficient, and trust must be made verifiable through institutional architecture. The social compact imperative argues that a society extending productive lifespans to 80, 90, and beyond while simultaneously disrupting careers through AI requires a new architecture of lifelong learning — not a patch on existing programs, but an investment of GI Bill ambition.

Running through all four arguments is a foundational distinction: AI excels at descriptive labor — pattern recognition, data synthesis, classification, iteration at scale. Human beings remain irreplaceable in prescriptive labor — judgment, values, accountability, and the weighing of competing goods. The goal is to maximize the first while protecting the second.

KEY FINDINGS

- Human expertise remains strategically valuable even as AI improves — high-stakes systems need people who can understand, audit, and override automated outputs.
- The governance challenge is not one problem but four: skill atrophy, AI system failure, strategic dependence, and community insecurity — each requiring different remedies but sharing a common solution principle.
- Process intelligibility matters more than output legibility in critical domains. Understanding how a system reaches its conclusions is foundational to meaningful human oversight.
- Voluntary industry self-governance is structurally insufficient. Trust must be built through verifiable institutional architecture, not asserted through corporate commitments.
- The coming wave of AI-driven displacement will coincide with dramatic increases in human longevity, creating a multi-decade transition requiring lifelong learning infrastructure that does not yet exist.
- The neo-Luddite risk is real and is best addressed by proactive social investment, not by dismissing legitimate grievances.

POLICY RECOMMENDATIONS

This paper proposes twelve actions across four actors. Industry should fund a Human Cognitive Advancement Initiative ($2B annual charitable investment modeled on Carnegie's library program), develop AI psychological safety tools, and establish a nonprofit engagement arm for community-level implementation. Government should establish a Lifelong Learning Transformation Fund ($7B, approximately 10% of the Department of Education budget, reallocated — not added) and a Workforce Transition Program modeled on the GI Bill. A Community Disruption Fund ($5–10B) should address small towns and urban areas facing acute displacement. Cross-sector partnerships should build AI governance and auditing standards through NIST and establish a stakeholder engagement platform. All of this should be governed by an accountability architecture with public reporting, pre-specified adjustment triggers, and independent oversight.

> *This is a pro-AI position, a pro-industry position, a pro-constitutional position, and a pro-human position at the same time. It is also the only position that takes seriously both what AI could accomplish and what is required — in governance, in social investment, and in human development — to ensure that it does.*

I. THE HOLLOWING: HOW AMERICA LOST ITS INDUSTRIAL MIND

From the 1970s onward, American firms increasingly moved production abroad in search of lower costs. The assumption was that the United States could keep design, engineering, and innovation while other countries handled the physical work. That assumption proved disastrously incomplete.

The overlooked factor was tacit knowledge — the embodied, experience-based judgment that cannot be fully codified in manuals, patents, or algorithms. As philosopher Michael Polanyi observed, we know more

than we can tell. When factories closed or moved abroad, the feedback loop between engineers and production floors was severed. Process and product innovation, once intertwined, drifted apart. American firms lost not only manufacturing capability but also the engineering depth needed to push the technological frontier.

There is an important distinction between two types of technological innovation: transformative breakthroughs that create entirely new industries, and the incremental, adjacent improvements that deepen and extend a technology's capabilities. The United States has consistently excelled at the first. But the second — small innovation — is equally important and often more durable, creating competitive moats that compound over time. When the underlying engineering base moves elsewhere, the capacity for small innovation follows.

The semiconductor industry provides the starkest illustration. By separating chip design from fabrication and transferring the latter to Taiwan and South Korea, U.S. companies assumed they could retain the "brains" while outsourcing the "hands." Two decades later, manufacturing excellence proved inseparable from engineering excellence. When supply-chain disruptions hit during the COVID-19 pandemic, the United States found itself strategically dependent on foreign production of the most critical component of modern technology. The CHIPS and Science Act — committing $52.7 billion to rebuild domestic semiconductor capacity — stands as an expensive admission that preserving ecosystems is far cheaper than reconstructing them.

> *Once capability disappears, restoring it is slow, expensive, and sometimes impossible. The same lesson applies to cognition.*

Industrial ecosystems behave like biological ones: remove a keystone species, and the surrounding system degrades. Knowledge is not a static stock; it is a living practice sustained by use. Lose the practice, lose the knowledge. Lose the knowledge, lose the ability to govern and improve the systems built upon it. America now stands at risk of repeating this mistake — not in physical production, but in the cognitive domains that underpin national power. That is the core meaning of intellectual protectionism.

II. AUTOMATION COMPLACENCY: WHEN SKILLS ATROPHY

High-reliability industries already understand the danger of over-automation. The foundational academic literature on human-automation interaction, established by Parasuraman and Riley in their landmark 1997 paper in Human Factors, identifies four failure modes: use, misuse, disuse, and abuse. The most consequential for the age of AI may be misuse causing humans to disuse their own capabilities — the automation complacency that develops when systems are reliable enough that operators lose the habit of independent verification. Parasuraman and Riley found that complacency occurs even in expert practitioners and cannot be overcome through simple training: the design of the human-machine system itself must preserve active human engagement.

On June 1, 2009, Air France Flight 447 departed Rio de Janeiro for Paris carrying 228 passengers and crew. It never arrived. The Airbus A330 was mechanically sound. The crew was experienced. The disaster stemmed from skill atrophy induced by over-reliance on automation. When pitot tubes iced over and the autopilot disconnected, the pilots — long accustomed to highly reliable fly-by-wire systems — struggled to diagnose inconsistent airspeed readings and recover from a stall. Manual flying skills had degraded through disuse; on long-haul routes, pilots might spend fewer than ten minutes per flight in

manual control. The aviation industry responded by strengthening manual-flight training requirements — the direct model for the Pilot-Hours Principle this paper recommends.

A parallel example appeared in naval operations. As GPS navigation became universal, the U.S. Navy discovered that officers were losing the ability to perform celestial navigation and manual charting. Naval academies reintroduced celestial navigation training in 2015 — a recognition that skill preservation is a national security requirement, not a nostalgic preference.

> *The more reliable automated systems become, the less frequently humans practice the skills needed to intervene when those systems fail.*

As AI systems assume more routine reasoning, drafting, classification, and synthesis work, human operators face the same dynamic across security, law, medicine, intelligence, defense, engineering, and finance. The ability to recognize an AI error can matter as much as the ability to produce a fast answer. This is why process intelligibility — understanding how a system reaches its conclusions, not merely whether its conclusions are correct — is foundational to meaningful human oversight.

III. THE CASE FOR COGNITIVE RESILIENCE

This paper does not argue for rejecting AI but for maximizing its value while managing its risks. Anthropic CEO Dario Amodei, in his 2024 essay "Machines of Loving Grace," offers one of the most serious cases for AI's transformative potential: a world in which AI compresses decades of biomedical progress into years, expands access to expertise that is currently scarce, and addresses problems that have resisted human effort for generations. The Stanford HAI AI Index Report 2025 documents the pace: AI benchmark performance advanced by 49 to 67 percentage points in a single year, global private investment hit a record $252 billion, and AI-enabled medical devices approved by the FDA grew from 6 in 2015 to 223 by 2023. That potential is worth pursuing. The question this paper asks is whether the institutional and human foundations required to govern, correct, and steer such a technology are being built alongside it.

The governance challenge has three distinct parts.

The first is human cognitive atrophy: if people no longer practice core reasoning tasks, they lose the capacity to supervise the systems performing them. The same Stanford report documents a 56 percent rise in reported AI-related incidents in a single year — 233 cases, including deepfake harm events and AI systems implicated in a teenager's death.

The second is AI system failure: models can hallucinate, behave unpredictably under edge conditions, or be manipulated through adversarial inputs. In an era of hybrid warfare, AI infrastructure is itself a target — as much as the military capabilities it enables. For civilian purposes, AI failure in health care, utilities, or financial services — where reliability must approach 100 percent — could be catastrophic.

The third is strategic dependence: when essential AI services concentrate in a small number of providers, societies become vulnerable to outages, capture, or coercive leverage. The Anthropic-Department of War dispute illustrates the dynamic in both directions — the Department could not operate with AI support

that might be withdrawn in critical moments, while the federal government's buying power creates pressure that could be used to compel providers toward surveillance or other civil liberties concerns.

These problems share a common solution principle: preserve human sovereign capacity. AI excels at descriptive labor — pattern recognition, data synthesis, classification. Human beings remain essential in prescriptive labor — judgment, values, accountability, and decisions for which someone must answer. The boundary is not always fixed: at sufficient scale, descriptive outputs shape prescriptive decisions, which is precisely why the boundary must be actively managed. AI should inform decisions; it should not constitute them.

IV. STRATEGIC DEPTH IN HUMAN INTELLIGENCE

National resilience depends on strategic depth: the ability to absorb shocks and continue operating under degraded conditions. The United States maintains this depth in military forces, energy systems, and industrial supply chains through redundancy, reserves, and domestic capability. The NIST AI Risk Management Framework, released in 2023 and widely adopted as the U.S. government's reference standard, explicitly identifies de-skilling and over-reliance as human-AI interaction risks that organizations must actively manage. Intellectual protectionism is the strategic complement to that framework: where NIST focuses on the AI system itself, this paper focuses on the human reserve that must exist alongside it.

A government that relies on AI for legal interpretation, intelligence analysis, financial risk assessment, or critical policy formulation must still be able to explain, review, and override those systems. Otherwise, public authority is being exercised through tools not fully transparent to the public or accountable to law. Democratic self-government requires that consequential decisions remain reviewable by human institutions that can assign responsibility. AI can assist those institutions, but it should not become the hidden source of binding authority where liberty, property, due process, or public safety are at stake.

> *Maintaining strategic depth in human intelligence is a constitutional imperative, not merely a technical preference.*

This framing is compatible with ambitious AI adoption. It does not require banning AI from critical domains. It requires building enough human capability, documentation, and fallback capacity that AI remains a tool under law rather than a system above it. The more consequential the decision, the more important it is that the people doing the reviewing have not had their competence hollowed out by years of delegation to machines they no longer fully understand.

V. THE TRUST PROBLEM AND ITS SUBSTITUTES

Public concern about AI is real and persistent. A 2025 Pew Research Center survey found that 50 percent of U.S. adults are more concerned than excited about AI; only 10 percent are more excited than concerned. Globally, the median is 34 percent more concerned than excited. Trust cannot be assumed; it has to be earned.

The trust problem is structural, not merely perceptual. The leading AI developers are simultaneously the entities warning about AI risk and the entities building the systems that create it, while their founders accumulate unprecedented wealth and influence. A company that issues a safety framework while retaining unilateral authority to override that framework's own recommendations has not solved the trust problem; it has documented it. OpenAI's Preparedness Framework Version 2 (April 2025) requests systematic evaluation of only 3 of the 24 risk categories identified in the academic literature, permits deployment of systems with "medium" capability for what OpenAI itself defines as severe harm (more than 1,000 deaths or $100 billion in damages), and concentrates final authority in the CEO. An independent analysis found the framework "does not guarantee any AI risk mitigation practices."

> *We cannot make AI leaders trustworthy by asking them to be. We can build institutional architecture that makes trustworthiness verifiable.*

There are three possible responses to the trust deficit. Authentic trust requires demonstrated willingness to absorb real costs for stated principles. When Anthropic refused to remove Claude's ethical constraints from a Pentagon contract — resulting in the Trump administration declaring Anthropic a "supply chain risk" — the action was credible precisely because it was costly. That is what genuine commitment looks like.

Structural trust requires governance architecture that does not depend on the personal integrity of founders: third-party auditing bodies with genuine authority, liability frameworks that expose companies to legal consequences for harms, and regulatory structures modeled on the FDA and FAA, where safety is an enforced standard with consequences for failure. We do not trust pharmaceutical products because their executives have good values; we trust them because they have been subjected to independent review by parties with no financial interest in the outcome.

Trust substitutes are what governance provides while authentic and structural trust are being built. They include disclosure requirements, transparency about known failure modes, human-readable audit logs, mandatory red-teaming before deployment, and the AI auditor profession this paper advocates. The AI industry is at the same juncture that aviation, pharmaceuticals, and finance reached before they earned genuine public trust. The frameworks for building that trust exist. The political will to require them is what is currently lacking.

VI. A POLICY AGENDA FOR COGNITIVE RESILIENCE

The policy response should be principle-based, not rule-based. AI changes quickly, and a static list of prohibited uses will age badly. Durable principles are better: preserve human autonomy, ensure meaningful review, require intelligibility where consequences are serious, and maintain the ability to recover when systems fail. The agenda below is organized across three streams, each targeting a distinct dimension of the governance challenge.

Stream One: Addressing Human Cognitive Atrophy

A domain classification system should distinguish contexts by consequence rather than by technology alone. Tier One includes domains where AI failure could be catastrophic, irreversible, time-critical, or constitutionally significant — nuclear command and control, biological threat assessment, systemic

financial risk management, constitutional legal interpretation. These require parallel human-run systems and routine AI-off drills. Tier Two includes consequential but recoverable domains: clinical medicine, intelligence analysis, military planning, policy research. Tier Three, where AI substitution is broadly benign, requires monitoring but not active intervention.

Selected professions should require periodic demonstration of AI-independent competence — the Pilot-Hours Principle. Medical licensing boards should require diagnostic reasoning without AI support as a condition of certification renewal. Intelligence analyst training should include substantial periods of unassisted analysis. Officer training at the service academies should devote significant curriculum to strategic reasoning without AI tools. Bar admission should test legal reasoning that cannot be delegated to AI-generated briefs.

Higher education and professional training should build process intelligibility rather than mere tool use. A new profession of AI auditors — domain experts trained to evaluate whether system outputs are trustworthy in context — should be developed for high-stakes domains. Industry and government should jointly fund the equivalent of a FASB standards body for General AI Application Protocols.

Stream Two: Addressing AI System Failure

High-stakes AI should undergo adversarial testing before deployment — red-teaming, sandboxing, and evaluation under degraded or manipulated conditions. Systems deployed in critical domains should provide human-readable documentation of known failure modes, confidence limits, and conditions of valid use. Critical institutions should maintain non-digital or air-gapped fallback procedures for essential functions. Where AI systems connect to sensitive networks, architectures should include quarantine, segmentation, and shutdown capability to prevent cascading failures.

Stream Three: Addressing Strategic Dependence

The legal framework should protect human autonomy, privacy, accountability, and review rights. Consequential AI decisions should remain subject to meaningful human oversight. Federal procurement is among the most powerful levers available: if the government requires auditability, human override, backup capability, and documentation as conditions of purchase, the market will build more resilient systems. Publicly funded knowledge — scientific research, government datasets, legal records, military doctrine — should be treated as a strategic commons governed with reciprocity, auditability, and security protections.

VII. THE SOCIAL COMPACT: SHARED OBLIGATIONS FOR A JUST TRANSITION

Governance frameworks address what institutions must do to make AI safe and accountable. They do not address the threat to livelihoods, communities, and life trajectories that AI deployment creates. That is a separate and equally important question — and the one most likely to determine whether the AI transition succeeds or fails politically.

The neo-Luddite risk is real and should not be dismissed as technophobia. The original Luddites were not anti-technology in the abstract; they were workers experiencing the costs of industrial transformation

without access to its benefits. The AI equivalent is already forming: Hollywood writers and actors won contract provisions limiting AI in production; Amazon and Google employees staged walkouts over AI military contracts; creative professionals are weaponizing copyright litigation against AI training data; advocacy organizations are mobilizing resentment over the concentration of AI's economic gains in a small number of firms and geographies. If these grievances are not addressed through proactive investment, they will be addressed through political disruption — moratoria, prohibitive regulation, or the kind of backlash that set back nuclear power for a generation.

A less-discussed structural factor makes the social compact more urgent than in any previous technological transition: the coincidence of AI disruption with the extension of human longevity. Medical advances are extending productive lifespans toward 80, 90, and beyond. A worker disrupted at 45 may need three or four complete professional reinventions over a 60-year remaining career. The conventional workforce transition program — designed to retrain a 52-year-old for a decade before retirement — is structurally insufficient for this reality. What is required is not a transition program but a new architecture of lifelong learning: education treated not as a phase of life that ends at 22 but as a continuous capacity that must be supported and affordable throughout a working life that may span six decades.

> *Andrew Carnegie, having built a fortune on the labor of industrial transformation, invested that wealth in the institutions that equipped a new generation for a new economy. The AI industry faces the same historical moment.*

Carnegie did not build libraries out of guilt; he built them because he understood that industrial wealth was built on a workforce whose capacities needed to be continuously renewed, and that investing in shared human capital served his industry's long-term interests. The social compact proposed here is not a tax or a penalty. It is the modern version of Carnegie's program — and the condition for sustained public legitimacy that commercial success at scale requires.

The nonprofit sector is the critical missing link in most AI governance discussions. Nonprofits — including faith-based organizations — are the connective tissue between national policy and community reality. A charitable investment that deliberately funds mediating organizations gives the social compact something no government program can manufacture: locally legitimate institutions that meet communities where they are, adapt programs to local conditions, and generate the bottom-up feedback that distinguishes what works from what looks good in a policy brief. This is the model that made Carnegie's library program work: not federal administration but local institutions, nationally funded.

The table below summarizes the social compact framework across four actors. The investments are large but not implausible: the Lifelong Learning Transformation Fund represents 10 percent of the Department of Education's current budget, reallocated rather than added. The Community Disruption Fund is comparable to a single medium-sized federal infrastructure program. The industry charitable investment, distributed across major AI developers, represents a fraction of their annual revenue growth. The question is not whether the resources exist; it is whether the political will exists to direct them before backlash forces a less orderly response.

ACTOR	INITIATIVE	MECHANISM & INVESTMENT
INDUSTRY	**Human Cognitive Advancement Initiative**	Consortium-funded philanthropic model. Minimum $2B annual investment — the modern Carnegie program — channeled through mediating nonprofits

		that design, deliver, and improve community-level programs. Targets lifelong learning infrastructure for all ages.
	AI Psychological Safety Tools	A Stay-Safe-Online analogue for AI: tools, curricula, and community resources grounded in developmental psychology. Helps individuals and families navigate AI from childhood through older age.
	Nonprofit & Community Engagement Arm	Dedicated funding stream for mediating organizations to build local trust, adapt programs to local conditions, and generate bottom-up feedback that improves what works.
GOV'T OR INDUSTRY	**Community Disruption Fund**	$5–10B targeted investment in small towns and blighted urban areas facing acute AI-driven economic disruption. Supports retraining centers, small business transition, civic technology access, and community anchor institutions.
GOVERNMENT	**Workforce Transition Program**	Federal retraining benefits, portable credentials, and income support for workers displaced by AI automation. Triggered by documented displacement events, not discretionary applications. Modeled on the GI Bill.
CROSS-SECTOR	**Lifelong Learning Transformation Initiative**	$7B annual investment — approximately 10% of the Department of Education budget, reallocated. Targets: guidance counseling for all ages; learning R&D and brain plasticity research; expanded course diversity in high schools; reinvented community college and university capacity for 30–80-year-old learners.
	AI Governance & Auditing Standards	NIST-anchored public-private standards partnership. Industry funds implementation; government sets principles; independent body certifies compliance. Modeled on aviation, pharmaceutical, and financial oversight architectures.
	Stakeholder & Community Engagement Platform	Cross-sector engagement infrastructure connecting AI developers, regulators, educators, employers, and community organizations. Generates the local legitimacy and feedback loops that national policy cannot manufacture from the top down.
ACCOUNTABILITY	**Metrics, Reporting & Adjustment Triggers**	Annual public reporting on displacement rates, retraining outcomes, community economic indicators, and AI incident data. Pre-specified triggers: if displacement outpaces transition capacity by more than 15%, mandatory fund increases activate. Overseen by GAO with independent academic review.

VIII. LEVERAGING EXISTING INSTITUTIONS

The United States does not need to invent an entirely new bureaucracy to begin. The NIST AI Risk Management Framework provides an existing, broadly adopted governance architecture. NIST can extend that work to develop standards for AI evaluation, auditability, documentation, and process-level intelligibility; its 2026 concept note on AI in critical infrastructure is a direct step in this direction. Sector regulators can adapt those standards to their domains: the FDA for medical AI, the SEC for financial systems, the FAA for aviation, DoD for defense applications. Procurement offices can make resilience a purchasing requirement. Courts and legislatures can clarify accountability and human review rights.

DARPA and the national laboratories can conduct adversarial testing and red-team exercises. The GAO and Inspectors General can ensure accountability for federal AI systems and flag cognitive dependency risks before they become irreversible. This distributed governance model avoids the risk of a single centralized AI regulator becoming a bottleneck or a capture target. Each institution brings domain authority and existing relationships with the professionals whose competence must be preserved.

The Lifelong Learning Transformation Initiative is essential for reorganizing the U.S. higher education and workforce development system around a fundamentally different mission. A system designed for 18-to-22-year-olds moving through education once is becoming increasingly obsolete. The future will require a system that serves 18-to-80-year-olds in continuous transition. Community colleges — already the most nimble and accessible institutions in higher education — are the natural delivery mechanism. Employers, unions, and professional associations are the natural co-designers.

A national convening on human agency and AI — bringing together AI developers, sector regulators, educators, community organizations, and civil society — could help align these actors around shared implementation priorities. When reports on educational decline emerged in the 1980s, President George H.W. Bush convened an Education Summit in 1989 that produced a generation of reform. The infrastructure for trustworthy AI should be built with the same seriousness that earlier generations applied to aviation safety, pharmaceutical oversight, and cybersecurity. Those industries are not trustworthy because their leaders chose honesty; they are trustworthy because institutions were built that made honesty verifiable and dishonesty costly.

IX. DESIGNING FOR HUMAN–AI COMPLEMENTARITY

The preceding sections address the risks that AI governance must mitigate and the social obligations it must fulfill. But a durable framework must also articulate what good AI deployment looks like — not as a constraint on technology, but as the standard by which its success should be measured.

Artificial intelligence is best understood as a productivity technology: one that expands the scope, speed, and scale of what human beings can do. The central design challenge is not whether to use AI but how to structure its use so that it consistently elevates human decision-making rather than displacing it.

Division of Labor: From Substitution to Elevation

AI's comparative advantage lies in pattern recognition across large datasets, rapid synthesis of information, simulation and iteration at scale, and consistency in routine tasks. These capabilities are most valuable when they allow human actors to move up the value chain of cognition — away from tasks that can be systematized and toward the tasks that require judgment, context, and accountability. AI is well suited to descriptive labor: data processing, classification, and pattern identification. Human beings

remain essential in prescriptive labor: judgment, values, trade-offs among competing goods, and decisions for which someone must be accountable. The objective is not to prevent AI from participating in decision-making but to ensure that final authority, especially in consequential domains, remains legible, reviewable, and human.

Decision Architecture: Supporting Judgment Under Uncertainty

Real-world decision-making is shaped by incomplete data, time pressure, competing objectives, and uncertainty about downstream effects. AI systems should be designed to support human judgment by surfacing uncertainty and confidence levels, making underlying assumptions visible, presenting alternative scenarios, and enabling users to weigh trade-offs. The goal should be to structure discretion — to give decision-makers the information landscape they need to act wisely, rather than presenting AI outputs as authoritative conclusions. Decision-makers should be empowered to operate along scales and ranges rather than binary rules, because risk tolerance varies by context, speed is sometimes properly prioritized over completeness, and least-worst outcomes are often preferable to theoretically optimal ones that cannot be achieved.

Skills Preservation: Capability as a Strategic Asset

Preserving human capability is not only a defensive measure against AI failure; it is a source of long-term competitive and democratic strength. AI can become one of the most powerful tools for documenting, transmitting, and extending human knowledge across time and communities — preserving endangered languages, capturing tacit knowledge from declining industries, recording expert practices in medicine, engineering, and craft. But skill preservation requires more than archival function: humans must retain the ability to perform and understand the tasks that AI supports. Effective knowledge systems should operate on two levels simultaneously — expert-level documentation that enables specialists to audit and refine work, and lay-level documentation accessible to non-specialists and decision-makers. A system that cannot be understood outside a narrow technical class is difficult to oversee; a system that lacks expert grounding is difficult to trust.

Education and the Expansion of Human Potential

The rise of AI does not reduce the need for human education. It increases it. If AI is deployed primarily as a substitute for human effort, educational incentives weaken over time; if deployed as a complement to human capability, the opposite occurs — education becomes more valuable as the returns to judgment, ethics, creativity, and contextual reasoning rise relative to tasks that machines assume. A healthy AI ecosystem must support a wide range of human aspirations. There are communities that seek to preserve traditional ways of life — the Amish model of selective technology adoption is a legitimate exercise of autonomy — and there are societies pursuing the most ambitious frontiers of science and exploration. These are not mutually exclusive. The role of AI should be to expand the feasible set of human outcomes, not to collapse it into a single model of efficiency.

Education policy in an AI age must hold two things simultaneously: rigorous investment in the technical literacy needed to understand and audit AI systems, and equally rigorous investment in the humanistic capacities — ethical reasoning, historical judgment, civic knowledge, creative and critical thinking — that AI cannot replace and that democratic self-governance cannot function without.

Complementarity as the Design Standard

The long-term success of AI will depend less on its raw capability than on how well it is integrated into human systems. A complementarity-based approach provides a clear and testable standard: does this system increase human capability, or does it obscure it? Does it support judgment, or does it replace it? Does it preserve optionality, or does it create dependency? These are not rhetorical questions; they are measurable design criteria that can inform procurement decisions, professional standards, regulatory requirements, and institutional design. Designing for human-AI complementarity is not an additional objective layered on top of the governance framework. It is the condition under which all the other objectives — economic, scientific, constitutional, and national — can be sustained.

X. THE STAKES

The point of this framework is to make AI more usable, not less. If the United States wants AI to scale across the economy and the institutions of governance, it has to solve the trust problem that comes with powerful systems operating in high-stakes domains. It also has to solve the legitimacy problem: a technology whose benefits accrue primarily to a small number of wealthy developers and early adopters, while its costs fall on workers, communities, and institutions unprepared for the transition, will not maintain the political conditions necessary for its own continued development.

The manufacturing parallel is instructive. America walked into deindustrialization with open eyes because short-term efficiency gains outweighed the institutional voice warning about long-term fragility. Reversing this has been a bipartisan priority for the past decade. Intellectual protectionism asks that the lesson be applied prospectively — to cognitive capacity, to the human reserve that AI governance requires, and to the communities that will bear the transitional costs of a technology they did not choose and cannot easily avoid.

Those who believe AI's risks are so severe it should be halted are wrong: the technology's potential to address genuine human needs is real, and the right response is not to stop but to build — the auditing professions, the testing requirements, the trust substitutes, the human reserve, the social compact, and the institutional architecture that make powerful AI governable.

> *The future of AI should be one in which machines do more of the routine work, while humans keep the authority, judgment, and institutional strength to direct what matters.*

That is a pro-AI position, a pro-industry position, and a pro-constitutional position at the same time. It is also the only position that takes seriously both what AI could accomplish and what is required — in governance, in social investment, and in human development — to ensure that it does.

ABOUT THE AUTHOR

Stephen C. Jordan is the Executive Director of the Institute for Sustainable Development, a practitioner-led think tank working on the conditions that determine whether communities can be built, sustained, and rebuilt. ISD's work spans original research, educational programming through the ISD Academy of

Fellows, and direct field engagement in disaster-affected and economically distressed communities. Website: isdus.org | Contact: sjordan@isdus.org

REFERENCES

1. Amodei, Dario. "Machines of Loving Grace: How AI Could Transform the World for the Better." darioamodei.com, October 2024. https://www.darioamodei.com/essay/machines-of-loving-grace

2. Andreessen, Marc. "The Techno-Optimist Manifesto." Andreessen Horowitz, October 2023. https://a16z.com/the-techno-optimist-manifesto/

3. Bureau d'Enquêtes et d'Analyses (BEA). Final Report on the Accident on 1 June 2009 to Airbus A330-203 Air France Flight AF 447. July 2012. https://www.bea.aero

4. Garreau, Joel. Radical Evolution: The Promise and Peril of Enhancing Our Minds, Our Bodies — and What It Means to Be Human. Doubleday, 2005.

5. National Institute of Standards and Technology (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. January 2023. https://doi.org/10.6028/NIST.AI.100-1

6. OpenAI. "Preparedness Framework Version 2." April 2025. https://cdn.openai.com/pdf/preparedness-framework-v2.pdf

7. Parasuraman, Raja, and Victor Riley. "Humans and Automation: Use, Misuse, Disuse, Abuse." Human Factors 39, no. 2 (1997): 230–253. https://doi.org/10.1518/001872097778543886

8. Pew Research Center. "How Americans View Artificial Intelligence and Its Impact on People and Society." November 2024. https://www.pewresearch.org

9. Stanford Institute for Human-Centered Artificial Intelligence (HAI). AI Index Report 2025. Stanford University, April 2025. https://hai.stanford.edu/ai-index/2025-ai-index-report

10. Thiel, Peter. Zero to One: Notes on Startups, or How to Build the Future. Crown Business, 2014.

Suggested Citation
Jordan, Stephen C. "Human Agency in the Age of AI: A Strategic Framework for Cognitive Resilience and Human Flourishing." Institute for Sustainable Development Policy Paper, 2026. isdus.org