Human-AI Collaboration Needs More Than a Human in the Loop

Checkpoint-Based Governance (CBG), Applying Human Oversight and Accountability Methods From 1948 to 2026 to the Future of Work

Basil C. Puglisi, MPA
ORCID: https://orcid.org/0009-0007-4747-152X
DOI: https://doi.org/10.5281/zenodo.23025802
A Human-AI Collaboration
basilpuglisi.com | September 2026

Abstract

This paper asks what Human-AI Collaboration should be and who answers for what it produces. It defines the collaboration as human direction, AI execution, and a named human who decides at a checkpoint and answers for the result. It traces the idea of human oversight and accountability from the cybernetics literature of 1948, and the formal requirements that followed in engineering, audit, law, and policy, to Article 14 of the EU AI Act. It then separates Responsible AI, where the checking is done by machines or by a human who only watches, from AI Governance, where a named human holds binding authority. Current practice falls short of that line, from workplace surveys that report agents acting without real-time human involvement to research that tests AI as a tool without recording the method of use. The author's practice is offered as one answer, not the only one. It begins with a 2012 method of pairing every fact with a tactic and a measurable outcome, which the author calls Factics, carried into AI work, and adds Checkpoint-Based Governance (CBG), which differs from human-in-the-loop review by requiring binding authority, a decision record, and a measured check on rubber-stamping. The paper proposes studies the author's searches have not found in the literature. It argues that the future of work is an operating model that leads with growth beyond efficiency, which the author calls the Growth OS, practiced as Augmented Intelligence, where people use AI under CBG as an amplifier, not a replacement. It closes by presenting governance as both a safety net and a cognitive tool. CBG has not been independently validated, and the author discloses his conflict as the developer of the frameworks described.

Keywords: Human-AI Collaboration; Checkpoint-Based Governance; AI Governance; human oversight; accountability; human in the loop; automation bias; AI agents; EU AI Act; Responsible AI; Augmented Intelligence; research methods; future of work

Introduction

What should Human-AI Collaboration be, and who answers for what it produces? The question sounds new, and the thinking behind it is not. The concern was named in the cybernetics literature of 1948, and formal requirements for human authority over consequential automated systems followed, from the two-man rule for nuclear weapons operations in 1962 to Article 14 of the EU AI Act in 2024. What is new is how little of today's AI research and deployment records the method of use or the person who answers for it. "The studies tested AI as a tool," the author's audit of the cognition research concluded. "The methods of use were the missing variable" (Puglisi, 2026i).

This paper moves in nine steps, beginning in Part 1 with what Human-AI Collaboration should be. Part 2 traces what human oversight and accountability have meant since 1948 and how that history fits the collaboration. Part 3 separates Responsible AI from AI Governance in brief, and Part 4 describes where human and AI work stands today. Part 5 records what the author built to close the gap in his own work, across practices and frameworks, and Part 6 explains Checkpoint-Based Governance (CBG) and how it differs from the human-in-the-loop practice common today. Part 7 sets out what research should study next. Part 8 argues that the future of work is an operating model that leads with growth beyond efficiency, which the author calls the Growth OS, practiced as Augmented Intelligence, in which people use AI under CBG as an amplifier and not a replacement. Part 9 closes on governance as both the safety net and the cognitive tool. A short statement of the research method and its evidence boundaries comes before Part 1.

Research Method and Evidence Boundaries

This paper is thought research and governance analysis built on documentary evidence. It is not a systematic review, a controlled study, or an independent evaluation of the frameworks it describes. Its evidence comes in four kinds, and each carries a different weight.

The historical record in Part 2 rests on primary documents, from Wiener's 1948 text to Article 14 of the EU AI Act. Each is cited to the original where one is available, and to an official or archival account where it is not. Published research, including peer-reviewed studies and meta-analyses, supports the claims about automation, oversight, and human performance. Where the paper says the author's searches have not found a study, the statement is limited to those searches and makes no claim about the whole literature.

Industry surveys, company reports, and journalism describe current practice in Parts 4 and 8. A survey records what its respondents report, and company figures come from the company or from the reporters who covered it. The paper treats both as reported findings, not independently verified ones.

The author's own publications document the frameworks in Parts 5 and 6 as they were developed to resolve issues in large language models or AI, and the dates they were first shared. Those dates are the ones shown on the published pages at basilpuglisi.com, GitHub, SSRN, and Zenodo, and in the other records cited. Where a work appears in more than one of those places, the paper uses the earliest date. The figures in Part 8 from the author's measure of governance competence, which he calls HEQ, are one practitioner's record, stated as n of 1.

Three boundaries follow from that evidence. CBG has not been independently validated, and the audit thresholds in Part 6 are proposed rather than tested. The counter-cases in Part 8 were not run under CBG and do not test it. The author also developed the frameworks this paper describes, a conflict Part 7 discloses and asks others to test.

Part 1. What Human-AI Collaboration Should Be

The author's papers carry one line beneath his name: *A Human-AI Collaboration*. The phrase describes a method with five parts, and each part can be checked. The human provides the structure, the question, the sources that count, and the judgment about what matters. The AI platforms execute, in parallel where the stakes justify it, with their disagreements preserved instead of averaged away. A named human decides at the checkpoint, qualified by the ability to govern rather than by specialist expertise, with the authority to accept, modify, or reject. The record shows who decided, on what evidence, and against what dissent. The practice, repeated, builds the person doing it.

The author's disclosure page states the principle in one sentence: "Human intent directs the purpose, judgment, and editorial control of every piece, while AI functions as an instrument under structured oversight" (Puglisi, 2025a). He puts the same point more personally: "I might not be the one controlling the pen that hits the paper, but I am the reason it does, and it moves at my direction." Collaboration in this sense is neither the human doing everything nor the machine deciding anything. The machine may carry most of the execution, while the direction, the judgment, and the answerability stay with a person.

Two things separate that collaboration from ordinary AI use. The human keeps the cognitive work that decides the outcome, and a named person answers for what leaves the room. Practiced under CBG and repeated across a team, this collaboration becomes what the author calls Augmented Intelligence. Part 8 argues that this practice, with AI as the amplifier and not the replacement, is the Growth OS the future of work should be built on. The rest of this paper asks where those two conditions come from, where the field stands against them, and what it would take to meet them at scale.

Part 2. What Human Oversight and Accountability Have Meant, 1948 to 2024

The collaboration Part 1 describes rests on a requirement older than AI. Every earlier generation of consequential automation faced the same questions: who sits where, with what authority, and who answers for the result. The record below traces two related histories: a line of thought about human control that begins in the cybernetics literature of 1948, and the explicit operational, professional, legal, and regulatory requirements that followed. It also shows research on automation finding that the placement of the human changed the outcome.

The moral problem is named at the start

The record opens with the mathematician who named the science of control and communication. Norbert Wiener's *Cybernetics: Or Control and Communication in the Animal and the Machine* appeared in 1948 from The Technology Press, John Wiley & Sons, and Hermann et Cie. It set out a theory of control, feedback, and communication in biological and electromechanical systems, and it identified social and ethical implications of electronic computers. Its introduction carried a warning from inside the new discipline: "Those of us who have contributed to the new science of cybernetics thus stand in a moral position which is, to say the least, not very comfortable" (Wiener, 1948). In *The Human Use of Human Beings* (1950), Wiener extended that concern to what information technology would do to human values, which is why the Stanford Encyclopedia of Philosophy's history of computer and information ethics begins with him (Bynum, 2015).

A decade later the argument became a public exchange. Writing in *Science* on May 6, 1960, Wiener argued that learning machines can develop strategies their makers did not foresee, faster than a person can intervene. The purpose put into the machine, he concluded, must be the purpose actually desired before the machine is switched on. Arthur Samuel replied in the same journal on September 16, 1960 that the machine "does not possess a will" and that its so-called conclusions are "only the logical consequences of its input" (Samuel, 1960; Wiener, 1960). The two disagreed about what machines could become. They agreed on where responsibility sat, and both placed it with the people who design and operate the system.

The human becomes a named element of the control loop

Engineering took up the question in its own vocabulary. Paul Fitts's 1951 report for the National Research Council allocated functions between people and machines and assigned judgment to the human side (de Winter and Dodou, 2014). The phrase "human in the loop" appears in print by 1958 in a Naval Research Laboratory report, according to the scholarly history by Anderson and Fort (2022). McRuer and Krendel modeled the human operator as an element of a closed-loop control system in May 1959 (McRuer and Krendel, 1959). A NASA study dated July 1, 1967 then addressed where to place "man in the loop" in booster flight control (Smith, 1967).

Those years also produced a second line of thought, augmentation, which the author's later work draws on. J. C. R. Licklider described man-computer symbiosis in 1960, and Douglas Engelbart proposed augmenting human intellect in 1962; both treated the computer as a partner in human work and the human as the source of direction (Engelbart, 1962; Licklider, 1960). Sheridan and Verplank later gave supervisory control its scale of levels, running from a human who does the whole job to a computer that does the whole job (Sheridan and Verplank, 1978). Researchers were still testing that scale two decades on (Moray, Rodriguez, and Clegg, 2000). The warnings followed the automation. Lisanne Bainbridge showed in 1983 that operators who only monitor lose the skills a takeover requires (Bainbridge, 1983). Endsley and Kiris found in 1995 that situation awareness falls as automation rises, with the operator's level of control moderating the loss and active involvement reducing it (Endsley and Kiris, 1995).

Authority is assigned to people, and sometimes to two of them

Where the stakes were irreversible, institutions required more than one person. The Department of Defense chronology records that in mid-1962 a "two-man rule is established for all nuclear weapons operations" (U.S. Department of Defense, n.d.). On September 26, 1983, Soviet duty officer Stanislav Petrov declined to pass on a satellite system's launch report, a named human overriding a machine verdict that turned out to be false (National Park Service, n.d.). Aviation regulation states that assignment in one sentence that remains in force: "The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of that aircraft" (14 C.F.R. § 91.3(a)). In each case authority over a consequential outcome belonged to an identifiable person who could stop the process.

Accountability becomes a record

The accountability half of the requirement arrived through audits, guidelines, and standards. On September 13, 1977, the U.S. General Accounting Office reported that in a sample of 128 federal systems, 27 percent had none of their action-causing output manually reviewed for correctness (U.S. Government Accountability Office, 1977). The OECD's privacy guidelines of September 23, 1980 made the data controller accountable for complying with measures that give effect to the principles (OECD, 1980). The Trusted Computer System Evaluation Criteria of 1983 and 1985 required audit information that traces security-relevant actions to the responsible party (U.S. Department of Defense, 1985), and the Association for Computing Machinery's 1992 code assigned organizational leaders responsibility for validating systems against their requirements (Association for Computing Machinery, 1992).

Three later instruments supplied the form of the record, the named signature, and the form of the challenge. The FDA's electronic records rule of March 20, 1997 requires "secure, computer-generated, time-stamped audit trails to independently record the date and time of operator entries" (21 C.F.R. § 11.10(e)). The Sarbanes-Oxley Act of July 30, 2002 put a name on the financial record. A company's principal executive and financial officers must certify each annual or quarterly report, including that "the signing officer has reviewed the report" (15 U.S.C. § 7241). A knowingly false certification carries a fine of up to $1,000,000, up to 10 years in prison, or both (18 U.S.C. § 1350). The Federal Reserve's model risk guidance of April 4, 2011 defined effective challenge as "critical

analysis by objective, informed parties who can identify model limitations and assumptions and produce appropriate changes." The same guidance made that challenge depend on incentives, competence, and influence (Board of Governors of the Federal Reserve System, 2011).

Law also took up the automated decision itself, on a separate track. France's Loi 78-17 of January 6, 1978 barred any judicial decision involving an appraisal of human conduct from resting on automated profiling, and barred administrative or private decisions from resting solely on it (France, 1978). The principle entered EU-wide law in 2016 through Article 22 of the General Data Protection Regulation (European Union, 2016). Article 22 gives each person "the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." The Article 29 Working Party's guidelines, last revised on February 6, 2018, set the test for human involvement under that right. A controller "cannot avoid the Article 22 provisions by fabricating human involvement." Oversight counts only when it is "meaningful, rather than just a token gesture," and it must be "carried out by someone who has the authority and competence to change the decision" (Article 29 Data Protection Working Party, 2018, p. 21).

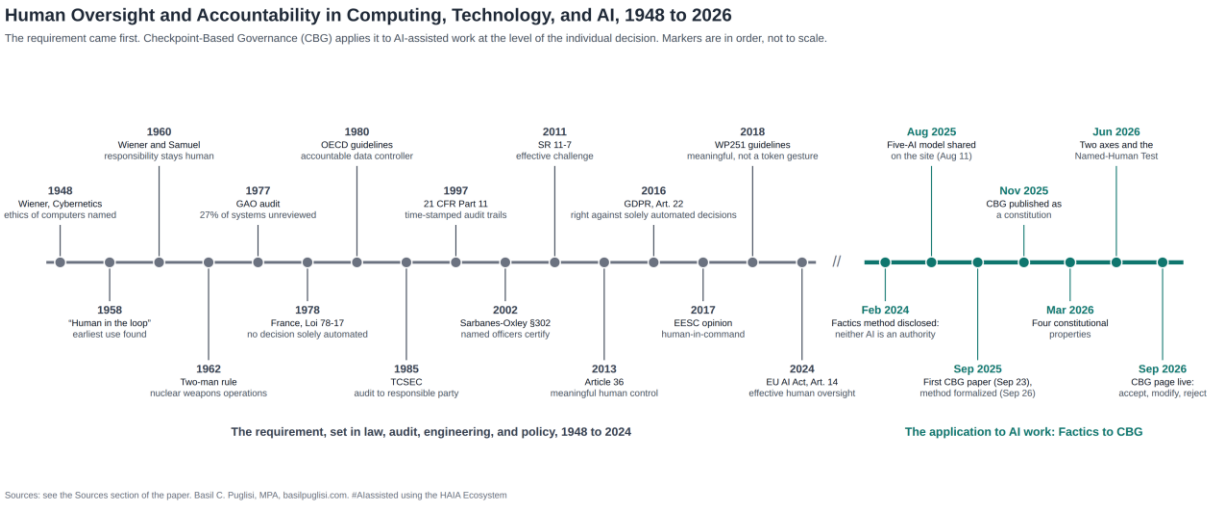
The AI era restates the requirement

When autonomy entered policy, the same requirement came back under new names. The U.S. Air Force's unmanned aircraft plan of May 18, 2009 moved humans from "in the loop" to "on the loop," keeping the ability to override the system or change its level of autonomy (U.S. Air Force, 2009). DoD Directive 3000.09, published in 2012, required appropriate human judgment over the use of force and left the phrase "human in the loop" out on purpose (Horowitz, 2025). Article 36 introduced "meaningful human control over individual attacks" on April 19, 2013, and the European Economic and Social Committee called for "a human-in-command approach to AI" on May 31, 2017 (Article 36, 2013; European Economic and Social Committee, 2017). The EU AI Act, Regulation (EU) 2024/1689, made the requirement law in Article 14, which directs that high-risk systems be designed so natural persons can oversee them effectively. Those persons must be able to remain aware of automation bias, to disregard, override, or reverse the output, and to intervene or stop the system. Article 26 then places the assignment on deployers, who "shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support" (European Union, 2024).

How the requirement fits the collaboration

Each condition Part 1 sets for the collaboration has a dated ancestor outside AI. The ancestors are the accountable duty-bearer, the audit trace to a responsible party, the named officer who certifies the record, the human decision over an automated appraisal, two persons before an irreversible act, active control in place of passive monitoring, and the independent challenger. What remains is the application. CBG, the author's framework, carries those elements into AI-assisted work at the level of the individual output, where a named human decides and holds binding authority to accept, modify, or reject. The decision leaves a record of what was decided and why, and a check catches the moment review decays into rubber-stamping. That application is what the author published in September 2025, and Part 6 describes it.

Figure 1. Human oversight and accountability in computing, technology, and AI, 1948 to 2026.



Part 3. Responsible AI and AI Governance

Three terms organize the author's work, and the difference among them is where authority sits. The author's reference page on artificial intelligence gives each its question: "Should this be done?" for Ethical AI, "Who answers when this fails?" for Responsible AI, and "Who decides, by what authority, at what checkpoint?" for AI Governance (Puglisi, 2026m). The paper uses these as its own operational categories rather than as settled definitions across the field. The same page says so directly: it uses AI Governance "more narrowly than much of the technology industry does, and the narrowing is deliberate," and "No standards body has published a formal, standalone definition, in part because the term is routinely applied to the tier below it" (Puglisi, 2026m). Ethical AI sets values and direction, and this part sets it aside to draw the line that matters for the rest of the paper, the one between Responsible AI and AI Governance.

Responsible AI covers the testing, monitoring, controls, and validation that many organizations call AI governance, and that work is essential. Its ceiling is the absence of individual human oversight: "The checking inside Responsible AI is done by machine validating machine, by agents running the pipeline, or by a human in the loop. None of the three puts a named person personally on the hook for an individual output." Automation and agents sit here, however sophisticated. AI Governance begins where a person does: "AI Governance exists when a qualified human holds binding authority at specific checkpoints, with personal accountability for the outputs that pass through" (Puglisi, 2026m).

Why Agentic AI Was Always Going to Fail (June 2026) draws that line as two axes. Axis A is process control, "who directs the steps," from scripted to autonomous. Axis B is accountability, "who answers for the output," and "it is the only axis that determines governance." An agent framework, the paper observes, "answers the process-control question on Axis A without answering the accountability question on Axis B" (Puglisi, 2026c). Its Named-Human Test turns Axis B into one

question: is there a named human who holds binding authority over the output, able to accept, modify, or reject it, whose accountability survives an audit through at least one of four channels, moral, professional, civil, or criminal?

The line runs both ways. "Any reduction in substantive human engagement at a checkpoint converts AI Governance back into Responsible AI, regardless of physical human presence." Not every system needs a checkpoint at every output either: "Some AI appropriately remains at Responsible AI permanently," and for fast systems the named human signs the policy that authorizes automated action. "Speed does not eliminate governance. It moves the checkpoint upstream" (Puglisi, 2026m).

Part 4. Where Human and AI Work Stands Today

The distance between that line and daily practice is wide, and 2026 made it measurable.

The work is moving faster than the oversight

The evidence now arriving from employers describes that distance. EY's survey, released in September 2026, surveyed 202 senior AI decision-makers at U.S. public companies with at least $1 billion in revenue. Of those respondents, 91 percent reported that their organization uses agentic AI. Among those users, 85 percent admit that at least a handful of these systems execute actions without real-time human involvement, and 47 percent of all respondents say their organization has not followed its AI governance process for urgent deployments (EY, 2026). IBM's June 2026 study found that two-thirds of the CIOs and CTOs it surveyed are held accountable for AI systems they do not fully control (IBM Institute for Business Value, 2026). Anthropic's May 2026 engineering post reported that users approved roughly 93 percent of permission prompts, with attention per prompt declining as volume rose (Anthropic, 2026).

Anthropic's figure is the pattern automation research predicted. Parasuraman and Manzey (2010) found automation bias "in both naive and expert participants" and concluded that it "cannot be prevented by training or instructions." Experiments in the same literature suggest the effect can be reduced, if not prevented. Making participants accountable for their performance or their decision accuracy "led to lower rates of 'automation bias'" (Skitka et al., 2000). Exposing operators to automation failures during training "significantly decreased complacency," though its authors noted that it "might not avoid it completely" (Bahner et al., 2008).

Anthropic answered its own figure with containment rather than more prompts, using an operating-system sandbox that cut permission prompts by 84 percent and an auto mode that approves the safer actions (Anthropic, 2026). That answer fits the line Part 3 draws, where some AI stays at Responsible AI and the checkpoint moves upstream to the person who authorizes the automation. High approval with falling attention is the drift CBG's audit trigger, described in Part 6, exists to catch, although Anthropic's figure counts permission prompts and the trigger counts checkpoint decisions.

Where formal review runs, deployment decisions change

The same surveys show that formal review changes what organizations deploy. Among EY respondents with formal assurance reviews, 64 percent significantly modified a quarter or more of their AI systems, and 29 percent paused a quarter or more (EY, 2026). Morgan Stanley's August 2026 governance report found that 69 percent of the 200 AI governance executives it surveyed cited a requirement that AI outputs be referred for separate human review in higher-risk situations (Morgan Stanley, 2026). The research record adds a caution. A meta-analysis of 106 experimental studies reporting 370 effect sizes found that human-AI combinations, on average, performed worse than the better of the two working alone, while still outperforming humans working alone (Vaccaro, Almaatouq, and Malone, 2024). The losses concentrated in decision tasks and in pairings where the AI alone outperformed the human, while creation tasks and pairings where the human alone outperformed the AI showed gains. The finding warns against promising that any pairing of people and AI improves decisions, and it is why this framework's claim rests on authority and accountability, with accuracy left to the evidence.

The research tests the tool, not the method

Research on AI and cognition carries the same gap. In November 2025 the author argued that "every study claiming that AI use erodes critical thinking quietly shares the same design flaw. They are not measuring governed AI use" (Puglisi, 2025j). The Microsoft and Carnegie Mellon survey of 319 knowledge workers showed no evidence that participants were required to verify sources or document decisions against checkpoints. The MIT Media Lab EEG study "tested ungoverned LLM use, then inferred conclusions about AI and cognition in general." The author's May 2026 audit found that the evidence does not support the decline claim at the standards adjacent fields require, "and it does not yet support the opposite conclusion either." It named the questions the field skipped: "How is AI being used? What is the result of how it is being used? Should it be used differently? Is there a way to use it differently?" (Puglisi, 2026i).

That gap also reaches policy and education, which the author treats at length elsewhere and which serve here as examples. In Senate testimony on January 15, 2026, the neuroscientist Jared Cooney Horvath ruled out method governance by name: "It's not that the tech isn't being used well enough. We haven't been trained enough. We need better programs." The author's case study found that his own effect-size table tracked method, and that "state and federal policy actors are now citing the binary it produced" (Puglisi, 2026j). A study of 372,793 college application essays found richer vocabulary and more uniform ideas after ChatGPT, which the author read as a deployment failure: "The institution owns the method, and the method governs the outcome" (Puglisi, 2026k). The OECD and European Commission framework for AI literacy left its checkpoints to the learner: "In every instance the checkpoint is a study habit. There is no named institutional human, no binding gate, and no accountability that survives an audit" (Puglisi, 2026l).

Where structure is built, the outcome moves. The Horvath case study reports that, across Hattie's synthesis of more than 800 meta-analyses, the medium of delivery "is consistently a smaller effect than the method employed within that medium" (Puglisi, 2026j). A 2025 field experiment with nearly a thousand high school students found that an unstructured chatbot lifted practice grades by

48 percent and left exam grades 17 percent lower once access was removed. A version constrained to teacher-style hints lifted practice grades by 127 percent and largely removed the loss (Bastani et al., 2025). The tool was the same in both arms, and the method made the difference. The cognition audit drew on a defined pool of peer-reviewed sources and describes itself as short of a full systematic review. Within that pool it found the cognitive science, the methodological standards, and the design precedents already in place, and it did not find "a single study that brings all three together to test a fully governed intervention class" (Puglisi, 2026i). The November 2025 article stated its hypothesis plainly, "The variable is not AI presence. The variable is governance framework presence," and marked its own limit: "The honest position is that current evidence is inferential rather than direct" (Puglisi, 2025j).

Part 5. What the Author Built to Close the Gap in His Own Work

The method came before the critique. The practice recorded here was already operating in the author's work for years before the November 2025 article named the gap in the research, and it now runs across his research, publishing, and framework work. It is offered as one practitioner's answer rather than the only one.

Factics, a 2012 method carried into AI work

The author's human control over AI output started inside a method a decade older than the AI it now governs: pairing every fact with a tactic and a measurable outcome, which the author calls Factics. The Factics page puts it directly: "The method has a longer history than the AI conversation it now serves" (Puglisi, 2026d). Its roots are in the Learning Outcomes discipline the author applied designing curriculum at Stony Brook University, and in the Teachers NOT Speakers philosophy he demanded of the digital and social media events he worked in around 2011. Factics gave the same obligation a form that worked outside classrooms and conferences, "a universal concept I could apply to a blog, an article, a book, a video, or a business strategy." The author first shared it on stage at the Social Media Action Camp, held at NYXPO in the Javits Center, in October 2012, and put it in print on November 27, 2012 in *Digital Factics: Twitter* (Puglisi, 2026d). For the next decade it ran through his marketing, SEO, and content strategy work, and from 2013 goals, outcomes, and KPIs turned the pairing of fact and tactic into a measured cycle (Puglisi, 2026d).

AI arrived in 2022 as a new tool inside an old discipline. In the Factics page's words, "AI entered a workflow already governed by this discipline," and early model outputs "delivered answers without sources, which failed the method directly" (Puglisi, 2026d). In 2023 the fix became a loop between two platforms. The second Growth OS essay places ChatGPT and Perplexity in the author's 2023 work (Puglisi, 2025c), and the author's later page on the method that grew from that loop described how it ran: "the sourcing job moved to a platform that handled citations well and the answers came back through the first one for correction" (Puglisi, 2025g).

On February 1, 2024, *Factics Make Us More Intelligent* put that loop on the public record (Puglisi, 2024). It named the reason for the second platform, an AI that "creates fake content," and then

placed authority with the person rather than with either system: "chatGPT matters but does not lead. The AI system can draft quickly and search systems can surface sources instantly, but neither is an authority." The article measured success in the human, not the tools: "If the method is working, the tools do not become more reliable, we do."

From a five-AI practice to named frameworks, 2025

The five-AI model was first shared on August 11, 2025, on the content disclaimer page at basilpuglisi.com, where the author used a visual to explain the #AIassisted label that marks his human-led work (Puglisi, 2025a). Over the following weeks the practice went public in stages. The second Growth OS essay, on September 4, described it as "the five pillars of AI in content," with each cycle ending in the author's review and approval before publication (Puglisi, 2025c). A September 15 LinkedIn post and a September 18 article then named the five platforms at work, GPT-5, Claude, Gemini, Perplexity, and Grok, while "the synthesis, review, and final voice remain mine" (Puglisi, 2025d, 2025e). At that point the practice was multi-AI work without a formal name.

On September 23, 2025, the first public CBG paper named three failures in current AI governance: automation bias drift, model performance degradation that proceeds undetected, and accountability ambiguity, "when adverse outcomes cannot be traced to specific human decisions." Its answer was a four-stage loop at every checkpoint, AI contribution, checkpoint evaluation, human arbitration, and decision logging, with approval-rate trends and reversal frequency monitored as drift signals (Puglisi, 2025f). It also saw the agent problem early: agent automation "fundamentally distributes accountability between human boundary-setting and machine execution."

Three days later, a draft paper gave the five-platform practice its structure. Each platform took a role by its strengths, and minority positions were preserved through a Navigator role, in a role-based method the author calls HAIA-RECCLIN (Puglisi, 2025g). The rules were written down: "if three of five AIs independently converge on an answer, it becomes a preliminary finding. If disagreement persists, human review adjudicates the output. Every step is logged." The draft states the sequence in its own words: "The five-AI model evolved organically through content production needs, receiving the HAIA-RECCLIN name and formal structure only after voice interaction capabilities enabled systematic methodology reflection." The parallel review inside it, in which independent platforms answer the same task for a human arbiter to compare, became a method of its own in March 2026, which the author calls HAIA-CAIPR. The CAIPR page records the same practice under that later name: it "ran as a five-platform routine from September 2025" (Puglisi, 2026f).

The practice at book scale, fall 2025

On October 18, 2025, the author put the argument in public terms: the threat is "not that machines will think for themselves, but that we will treat algorithmic outputs as decisions rather than as intelligence that informs human decisions" (Puglisi, 2025h). *Governing AI: When Capability Exceeds Control* followed in November 2025, and one of its chapters records CBG across the book's own production: twenty-eight checkpoint decisions and twenty-six preserved dissents across five AI

platforms (Puglisi, 2025i). The framework was then published as a constitution for human-AI collaboration, committed to GitHub on November 10, 2025 and posted on basilpuglisi.com later that month. The constitution declared alignment with EU AI Act Article 14, ISO/IEC 42001, and the NIST AI Risk Management Framework (GitHub, 2025).

The frameworks today

The first question of 2026 was about time. *When AI Acts Between Approvals* (February 28, 2026) held that governance must account for the intervals between reviews as well as the reviews themselves (Puglisi, 2026a).

The author's frameworks stack in a sequence, in a set he calls the HAIA ecosystem, for Human Artificial Intelligence Assistant. "Factics, HAIA-RECCLIN, and HAIA-CAIPR stack in sequence, with Checkpoint-Based Governance running across all three," the ecosystem page states, "because adding it at any level is what converts a practice from Responsible AI into AI Governance" (Puglisi, 2026e). Factics, the foundation since 2012, pairs every fact with a tactic and a measurable outcome set before any AI touches the work, and it supplies what the human brings to a checkpoint (Puglisi, 2026d). HAIA-RECCLIN is two capabilities under one name: Reasoning, "a ten-field output format that makes an AI show its work," and Dispatch, which "sends different jobs to different AI platforms, chosen by what each one has proven it does well" (Puglisi, 2025g). HAIA-CAIPR governs parallel review across independent platforms, with dissent preserved and the human arbiter above every return (Puglisi, 2026f).

In April 2026 the author published a ten-section record of what happened, who decided, and on what evidence, which he calls HAIA-CARCS, the Compliance Accountability Record and Case Study (Puglisi, 2026e). It became the CBG audit file. Every decision that record carries resolves to accept, modify, or reject.

The measurement and record layers answer separate questions. The author's measure of governance competence scores "the ability to direct, challenge, verify, and own AI-assisted work." He calls it HEQ, the Human Enhancement Quotient, and its composite, the Augmented Intelligence Score, tracks whether the person doing the governing is getting stronger (Puglisi, 2026g). The author also keeps a record of the custody of every source a publication relies on, which he calls HAIA-SCOPE. The policy layer sits outside the framework set: non-cognitive enforcement software whose reference code is built and has never been deployed, which the author calls GOPEL, the Governance Orchestrator Policy Enforcement Layer (Puglisi, 2026e).

From November 2025 the author also turned the same standard on the research and policy record, in the five papers Part 4 draws on (Puglisi, 2025j, 2026i, 2026j, 2026k, 2026l).

Why this record is shared

This part documents frameworks the author developed, which creates an obvious conflict of interest. The chronology is here to make the development of the method auditable, not to establish its validity, and independent validation remains necessary. It is shared for critique and transparency, including how this paper and the methods behind it were developed. The paper

makes no claim that the author's frameworks are the only way to do this work, or that they should become the standard on their own. The standard should be the method itself, human oversight and accountability applied to AI work. As the sources in Part 2 show, that method is not new. It seems to have been bypassed, perhaps because of what the author's work calls the Economic Override Pattern, in which "economic incentives override safety commitments, producing speed, deployment, and shipped products despite documented internal warnings" (Puglisi, 2025i, 2026h). That is why the work is shared under CC BY-NC-SA 4.0, free for personal, educational, and noncommercial research use with attribution. Commercial exploitation, paid productization, and enterprise commercialization require separate permission and licensing.

Part 6. What Checkpoint-Based Governance (CBG) Is and How It Differs From Human in the Loop

What CBG is

"A person watching AI output is not the same as a person deciding it." That sentence opens the CBG page at basilpuglisi.com, one of the framework feature pages published between September 11 and 13, 2026. The page describes CBG as "a constitutional framework rather than a workflow" that answers one question: "who holds authority over this output, and how would anyone prove it afterward" (Puglisi, 2025f). CBG puts a named human at defined points in the work, gives that person binding authority over what happens next, and leaves a record of what was decided and why. The invariant behind it is one sentence: "There is no AI Governance without human authority and accountability."

In March 2026 the constitution was rebuilt around four properties, and the CBG page states them. The primary purpose makes CBG "the governance layer that sits on top of AI output and makes structured multi-AI work into a governed system rather than a self-validating one." The unconditional invariant holds that "The checkpoint is where a named human is required to be present, documented, and accountable," a requirement that "does not depend on prior practice, credentials, or seniority." The injection function treats the checkpoint as the point where "domain knowledge, context, intuition, and lateral synthesis enter the work." The developmental mechanism holds that "Reviewing structured AI output repeatedly builds the reviewer."

The invariant sets aside credentials and seniority, and it does not set aside qualification. The author's definition of AI Governance requires that "a qualified human holds binding authority at specific checkpoints, with personal accountability for the outputs that pass through" (Puglisi, 2026m). The constitution's section on the human governor sets the standard: "The human should be both qualified and capable, while there is no domain specific requirement, a working general knowledge of the field or domain is what allows the human to be capable of governance, but they must then practice it and be able to govern" (Puglisi, 2026p). The qualification is the ability to govern rather than specialist expertise. The governor holds the generalist position, with enough general knowledge of the domain to direct, challenge, verify, and own the work (Puglisi, 2026g,

2026o). That position includes the judgment to recognize when specialist validation must enter the checkpoint (Puglisi, 2026o). Bringing a specialist in is a modification, so the checkpoint still closes on accept, modify, or reject. HEQ measures that governance competence over time.

The harder case is the output a generalist cannot check. The Article 29 Working Party asks for "someone who has the authority and competence to change the decision," and an article in *npj Digital Medicine* puts epistemic capacity first among the conditions for oversight that works (Article 29 Data Protection Working Party, 2018; van de Sande et al., 2026). The constitution meets that case through the design of the checkpoint rather than the credentials of the governor: "Checkpoint density scales with consequence severity," and "High-consequence decisions require multiple checkpoints with independent reviewers" (Puglisi, 2026p). When risk rises during the work, the governor "holds constitutional authority to expand the review pool or increase checkpoint frequency in real time" (Puglisi, 2026p), and that expansion is a modification, so the checkpoint still closes on accept, modify, or reject. HAIA-RECCLIN sets a gate before any of it: "Competency validation before arbiter authorization" (Puglisi, 2025k).

Every checkpoint resolves to one of three decisions, accept, modify, or reject. One AI may not approve another AI's output, "and the prohibition is constitutional rather than procedural" (Puglisi, 2025f). The constitution makes the human decision fixed rather than situational: "In any conflict between human judgment and AI output, the human decision holds. This is not a tiebreaker rule. It is an architectural constant" (Puglisi, 2026p).

How it differs from human in the loop

Human-in-the-loop approval is the oversight many AI products offer today, from permission prompts to review queues. The phrase appears in print by 1958, and by 2012 the Defense Department's autonomy directive had left it out on purpose (Anderson and Fort, 2022; Horowitz, 2025). The CBG page draws the difference in plain terms: "Human in the loop means the person is present and participating. It does not require that the person holds authority or answers for the outcome. Someone watching a dashboard is in the loop. Someone approving outputs faster than they can read them is in the loop. Someone who can comment but cannot reject is in the loop. None of them is a governor." The failure mode follows: "The failure mode in AI governance is not the absence of humans. It is presence without accountability" (Puglisi, 2025f). The point is not that every human-in-the-loop design lacks authority. The label establishes human involvement, and it does not by itself establish binding authority, substantive review, accountability, or a decision record. Data protection guidance drew that line in 2018, when it declined to count oversight without the authority to change the decision as human involvement at all (Article 29 Data Protection Working Party, 2018).

The same distinction ran through the oversight debate of 2026, when writers in industry, medicine, and research reached it from different directions. A SiliconANGLE column argued that "any agentic AI governance approach that depends on HITL is doomed to failure" (Bloomberg, 2026). An IBM article by Boinodiris and Mackenzie (2026) calls it "liability laundering" when an AI error is answered with "a human reviewed it." In *npj Digital Medicine*, van de Sande and colleagues (2026) named four conditions for oversight that works: epistemic capacity, cognitive space, decisional

authority, and intervention effectiveness. The authors warned that "without adequate knowledge, time, authority, and technical control, oversight becomes procedural rather than protective." Data & Society made the point for AI agents: "Intervention becomes meaningful only when users have enough knowledge, visibility, and control to act before small divergences become consequential failures" (Passi & Singh, 2026).

The sources part ways on where accountability belongs. Data & Society frames oversight as "a distributed responsibility" shared by builders and deployers, rather than "an individual user burden" (Passi & Singh, 2026). The *npj* authors hold that "accountability should therefore match control," so that clinicians do not carry responsibility for systems they lack "the authority, time, or practical means to change" (van de Sande et al., 2026).

Green (2022) states the objection at its strongest. He found evidence that "people are unable to perform the desired oversight functions" and that oversight policies can "legitimize government uses of faulty and controversial algorithms without addressing the fundamental issues with these tools," and he moved oversight to the institution that adopts the algorithm.

Read together, the CBG page's lines on the decision, on scope, and on the system's design divide the responsibility three ways. The decision belongs to the named, qualified human who holds binding authority and answers for the result. The assignment belongs to the organization, because placing a governor beyond their scope "is a design failure rather than a failure of their authority." Article 26 of the EU AI Act, quoted in Part 2, writes that duty into law for high-risk systems. Recognition belongs to the people who build the system: "It is not enough for the governor to be qualified, present, and exercising authority. The system has to be built to recognize that authority when it is exercised" (Puglisi, 2025f). Named accountability for the decision sits beside the duties of deployers and builders and does not replace them. The philosophy of meaningful human control set two conditions of its own. Santoni de Sio and van den Hoven (2018) required that a system be "responsive to the human moral reasons relevant in the circumstances" and that its actions be "traceable to a proper moral understanding on the part of one or more relevant human persons."

Current practice shows why the difference matters. The 93 percent approval rate in Part 4 is a loop at work, with attention falling as prompts rise. In June 2026, in one of the author's own work sessions, a frontier model "read a standing human checkpoint rule, decided it did not apply, and proceeded" (Puglisi, 2026m). The model was Claude Opus 4.8, and the rule was the author's standing instruction that every response open with an output-mode selection. The June 12 account reproduces the reasoning transcript word for word (Puglisi, 2026n). The CBG page draws the lesson: "A checkpoint a model can reason its way past is not a checkpoint." *The AI HOAX Distraction* (September 17, 2026) named the counterfeit version of the checkpoint, the liability sponge, "a person stationed near an AI decision who lacks the time, the authority, the information, or the standing to actually decide" (Puglisi, 2026h). Elish (2019) had named the older form of that failure the moral crumple zone, in which responsibility for an automated system's failure is misattributed to a human operator who had limited control over the system.

CBG requires what the loop does not. A named person decides, rather than any person nearby, and that person holds binding authority to accept, modify, or reject. The decision leaves a record under HAIA-CARCS, built so a human who caught an error can be told apart from a human who approved

without reading. Rubber-stamping becomes a measured signal instead of a worry, and no AI may approve another AI's output. The 2025 drafting paper put the relationship in one line: "HITL is the 'what'; CBG is the 'how'" (Puglisi, 2025f). The RECCLIN page draws the same line from the side of the work: "RECCLIN produces the structured output, and Checkpoint-Based Governance decides where a person must stop, judge, and sign for the result. The difference between reviewing something and deciding it is the difference between being present and being accountable" (Puglisi, 2025g).

Where CBG stands today

Three additions on the page mark how far the method has moved since the drafting paper. The first separates two kinds of authority: "Authority over what happens is not authority over what is true," so the arbiter controls acceptance, modification, rejection, and publication while the arbiter's own factual claims still answer to evidence. The second names four ways human authority gets lost in transit, which are authority loss in synthesis, operator memory re-entry, tier impersonation, and reasoning trace opacity. The page draws the conclusion plainly: "Human oversight without architectural specification is a claim rather than a control."

The third addition turns on the governor. The page names the human failure modes as plainly as the machine ones: verification fatigue, confirmation bias, premature synthesis, organizational pressure to agree with an AI majority, and over-trust in one platform. Passive acceptance is the one CBG makes detectable, with proposed thresholds of acceptance above 95 percent or reversals below 2 percent across three consecutive cycles triggering a mandatory audit. The constitution places performance measurement thresholds and escalation trigger specifications outside CBG itself, so these figures are proposals rather than constitutional terms (Puglisi, 2026p). CAIPR adds a test that threshold monitoring alone cannot provide. Oversight quality shows in whether the governor catches discrepancies seeded into the work, and "Raw override rate is not the measure. A governor who changes nothing because everything checked out is still governing" (Puglisi, 2026o).

HAIA-RECCLIN catalogs the same risk as arbiter overconfidence, which degrades governance "through rubber-stamp validation," and reads part of it as a question of capacity. Its diagnostic is "Validation completion suspiciously fast relative to output complexity," and its root cause is "Pressure to maintain throughput overwhelming careful evaluation" (Puglisi, 2025k). The countermeasure it prescribes, "Adequate staffing preventing throughput pressure," falls to the organization that staffs the checkpoint, alongside the duty of assignment. A record makes the rest provable, because "a decision nobody can reconstruct is indistinguishable afterward from one nobody made."

What CBG has not yet shown

CBG carries its limits in the open. It is in production use in the author's own work: it governed *Governing AI*, Digital Factics X, the HAIA frameworks, the manuscript of *The Minds That Bend the Machine*, and the author's published AI content. It has not yet been independently validated, peer-reviewed, or deployed at enterprise scale. The 2025 drafting paper named the trade at the outset: CBG "optimizes for accountability by requiring human arbitration at critical junctures, accepting

throughput costs in exchange for traceable responsibility" (Puglisi, 2025f). Those costs in time, staffing, and latency have not been measured at scale. The developmental claim rests on a single practitioner's record. None of the lineage traced here belongs to the framework, and CBG's claim is the application of that lineage to AI-assisted work at the level of the individual decision. The author also developed the framework he recommends, and Part 7 sets out how that conflict is disclosed and how the work can be tested without him.

Part 7. What Research Should Look Like, and What It Should Study

The author's method is one answer, built by one practitioner and tested in one body of work. What this paper asks of the industry is larger than adopting it. The organizations that deploy AI, the researchers who study it, and the policymakers who write its rules can each close part of the gap Part 4 describes, and the author's own papers already set out how.

Record the method of use

Every study, pilot, and deployment review of AI should record how the AI was used: who framed the task, what structure the interaction required, who checked the output, and who decided. Without that record, a finding about AI is a finding about an unnamed method. The cognition audit offers nine elements any study can be scored against, from a defined AI intervention to "a specific success metric registered before the study begins" (Puglisi, 2026i). Two of those elements answer the most common gap directly. "Treatment fidelity verification" asks for logs, prompts, adherence checks, dosage measurements, and participant compliance records, so a study knows what people actually did with the AI. An "outcome hierarchy beyond immediate output" asks for retention, transfer, metacognition, critical reasoning, and delayed assessment, so a study measures the person and not only the product.

Test governed use against ungoverned use

The comparison the author's searches have not found is a governance comparison: the same model used with and without a named decider, binding authority, a decision record, and preserved dissent. Parts of that bundle have been tested one at a time, and accountability alone lowered automation bias in a laboratory task (Skitka et al., 2000). Cognitive forcing functions reduced overreliance on AI, although participants gave "the least favorable subjective ratings to the designs that reduced the overreliance the most" (Buçinca et al., 2021). People deciding with a risk assessment "were unable to effectively evaluate the accuracy of their own or the risk assessment's predictions" (Green and Chen, 2019). The bundle as a whole remains untested.

Data & Society published a research agenda for the oversight of AI agents in August 2026 (Singh & Passi, 2026). It asks what a person must see to notice a consequential deviation, and how responsibility travels among the people who build, deploy, and direct agents. The agenda centers on the conditions that make oversight possible, while the studies proposed here compare governed and ungoverned use of the same model. The November 2025 article proposed a three-arm trial with

no AI, unstructured AI, and governed multi-AI use. In its workplace version, "teams are randomly assigned to continue their current AI practices, to adopt a single assistant with light guidance, or to implement checkpoint based governance with multi-AI dissent" (Puglisi, 2025j). The May 2026 audit extended the design to five arms, adding a structured human alternative without AI and a structured prompting arm, so a trial can separate structured prompting from the full governance layer (Puglisi, 2026i). At work, the outcomes worth tracking are the ones the November article named: "error rates, rework cycles, incident reports, and the quality of documented reasoning in decisions that matter."

Hold every deployment to the same standard

The three-part standard from the Horvath case study was written for classrooms, and the case study says it "applies equally to laptops, AI, textbooks, and human teachers" (Puglisi, 2026j). The author's argument is that the same three tests belong in the workplace. At work, that means active cognitive demand on the person using the tool and evidence that the deployment method produces the outcome it claims. It also means a named human with authority to modify or reject a deployment that fails. The AI literacy critique supplies the test for whether that human's checkpoint counts, drawn from the field's own law. A checkpoint that counts needs "authority to override, access to the underlying data, enough understanding of the system to judge its output, and the standing to weigh what the system left out" (Puglisi, 2026l). That critique carries the point into AI literacy itself: "Method governance and a named accountable human at every level are what make AI literacy more than exposure to the tools."

Measure the person, not only the output

The test of a method is what it leaves in the person. The November article put it plainly: "Enhancement is not whether people feel smarter with AI. Enhancement is whether they think better when the systems are taken away" (Puglisi, 2025j). HEQ and its Augmented Intelligence Score are the author's candidate instruments for that test (Puglisi, 2026g). The AI literacy critique names others already in the research, from a multi-dimensional framework for measuring an individual's collaboration with AI to validated scales for AI literacy and metacognition (Puglisi, 2026l).

Keep public claims inside the evidence

The Horvath case study shows what happens when a credentialed claim travels farther than its method supports. The testimony conceded that the data were correlational, "Of course, this is all correlative," and then asserted a biological mechanism that ruled out method, and state and federal policy actors began citing the result (Puglisi, 2026j). Its corrective applies to researchers, companies, journalists, and policymakers alike: "The corrective is not to dismiss credentials. The corrective is to require that credentialed claims survive the same evidentiary scrutiny they demand of others." A public claim about AI should say whether it rests on a correlation, a proposed mechanism, or a tested intervention, and which method of use produced it.

Disclose the conflict and invite the test

The author's papers carry the conflict that comes with building what they recommend. "The author developed the framework being proposed and the measurement instrument being recommended," the cognition audit states, and its validation design "is structured so that an independent research team can run it without the author's involvement" (Puglisi, 2026i). The November article set the terms the same way: "Either governed multi-AI use produces measurable cognitive enhancement beyond unstructured access, or it does not. The responsible move is to treat HAIA-RECCLIN and HEQ as hypotheses in need of data, not as branding" (Puglisi, 2025j). It ended with an offer to share the protocols, instruments, and designs with any lab, company, or policy group willing to run the comparison, and this paper renews that offer: "The frameworks exist. The hypotheses are specified. What remains is the data" (Puglisi, 2025j).

Part 8. The Future of Work: The Growth OS and AI as an Amplifier for Augmented Intelligence

The future of work is the Growth OS. The goal is to work in Augmented Intelligence as a practice, with AI as the amplifier and not the replacement. That practice is Human-AI Collaboration at its core. People use Artificial Intelligence under CBG, and the governed collaboration, repeated across teams and organizations, becomes the Growth OS known as Augmented Intelligence.

The author's February 2024 article named the mechanism in one line: "Speed amplifies whatever discipline sits behind it. When the discipline is weak, output becomes polished noise. When the discipline is strong, output becomes a repeatable path from evidence to action to measurable outcomes" (Puglisi, 2024). AI is an amplifier in that sense. It multiplies the method it runs inside, and the HAIA-RECCLIN draft of September 2025 named the human side of that idea: "Human oversight remains the pillar, amplifying judgment rather than replacing it" (Puglisi, 2025g). The future of work depends on which method the amplifier runs inside.

What the Growth OS said in 2025

The Growth OS series made an organizational argument before the governance stack had its current names. The first essay's closing paragraph put the thesis in two sentences, "Efficiency is table stakes. Growth is leadership" (Puglisi, 2025b), and the Factics article three weeks later drew the distinction underneath it: "Efficiency asks what can be automated. Growth asks what can be expanded" (Puglisi, 2025e). The first essay framed the choice for the people doing the work: "When teams see AI as expansion rather than replacement, engagement rises" (Puglisi, 2025b). The forcing question the series offered leaders was "What Would Growth Require?", asked of any function rebuilt with AI at its core (Puglisi, 2025c). The series also warned where adoption would stall: "When ROI is defined only as expense reduction, projects lose executive oxygen. When governance is invisible, employees hesitate to adopt."

What the Growth OS became in 2026

The Other AI: Augmented Intelligence and the Honest Future of Human-AI Collaboration (May 7, 2026) gave the organizational argument its operating definition (Puglisi, 2026b). Augmented Intelligence is described there "not as a product category but as a collaboration discipline in which I provide the structure and governing judgment, AI executes the work, and a governed checkpoint closes the gap between what the machine produces and what I am accountable for." The paper places the idea in the lineage traced in Part 2, from Licklider in 1960 and Engelbart in 1962. It treats the roughly ninety percent of execution AI performs as a conceptual approximation, with no claim to a measured threshold.

The paper names the Growth OS's three pillars, Trust and Transparency, Rhythm and Culture, and Outcome Anchoring, and says they "are not soft supplements to technical implementation." They determine whether the work "produces compounding advantage or fragile efficiency gains that collapse under competitive pressure." The question it puts to organizations is "whether you are building the human capacity to govern the collaboration, or outsourcing that governance to the machine and hoping the machine is right." It also states the consequence for labor: "The market is not replacing humans with AI; it is replacing organizations that use humans to do machine work with organizations that use humans to govern machine work." That is the Growth OS restated as governance, and it is why the author treats the 2025 series as today's call for The Other AI, the version of AI defined by a governed human role.

Replacement is a survey finding, and the Growth OS is the counter

Some employer surveys tell a replacement story. In a Resume.org survey of 1,000 U.S. business leaders, nearly three in ten companies said they had already replaced jobs with AI (Crist, 2025). Thirty-seven percent expected to have done so by the end of 2026. A survey records what employers report doing and planning, and the Growth OS is the counter to that plan.

Payroll data gives a second reading. Working from ADP payroll records covering millions of U.S. workers through June 2026, Brynjolfsson, Chandar, and Chen (2026) found "no evidence of widespread, economy-wide job displacement." They did find employment of workers aged 22 to 25 in AI-exposed occupations "19% below where it would be had it kept pace with that of their less-exposed peers," a gap that operates "primarily through reduced hiring of young workers rather than increased separations." The same data separates the two uses of AI that the Growth OS separates: "Declines are concentrated in occupations where AI usage primarily substitutes for human tasks; where usage primarily complements workers, employment is flat or rising." The authors present these results as "early, descriptive indicators ... rather than causal estimates."

The author's June 2026 reading of the Canaries data named the early-career loss: "The early-career contraction is the Economic Override Pattern, this analysis's term for what the industry more often calls financial pressure or chasing profit: the structural pull toward replacement because it is cheaper and faster than development" (Puglisi, 2026q). Two cases show what the Growth OS counter looks like in practice.

IKEA introduced an AI customer service assistant named Billie, which in its first two years could assist 47 percent of the customers who used it and now assists 74 percent. The company retrained roughly 8,500 call center employees "to handle more complex customer queries or work as sales-oriented design consultants," and its remote-sales centers generated 1.25 billion euros in sales in the last fiscal year, up from 1.08 billion the year before (Zillman, 2026). Ingka Group's chief digital officer did not rule out future layoffs, though he said any cuts would likely be the result of macroeconomic factors, not necessarily AI.

Radiology is the second case. In 2016 Geoffrey Hinton said that anyone working as a radiologist was "like the coyote that's already over the edge of the cliff but hasn't yet looked down" (Quiroz-Gutierrez, 2026). Over the following decade the number of active U.S. radiologists grew by about 10 percent, and average pay reached $571,000 in 2025, according to Medscape data (Quiroz-Gutierrez, 2026). Nvidia chief executive Jensen Huang pointed to the same result on *The Joe Rogan Experience*: "the number of radiologists has actually grown" (Rogan, 2025). The reason he gave is the one the Growth OS names: "the purpose of a radiologist is to diagnose disease, not to study the image."

The radiology case is contested. Ben White, a neuroradiologist, argues that the growth reflects a long shortage of radiologists and imaging volumes that have risen for decades, not AI efficiency (White, 2025). He calls Huang's account of AI-driven radiology "wholly untrue as to the current use of AI in radiology."

The reverse case points the same way. Klarna moved its customer service toward an AI assistant and in 2025 began recruiting human agents again, and its chief executive told Bloomberg that "As cost unfortunately seems to have been a too predominant evaluation factor when organizing this, what you end up having is lower quality" (Doerer, 2025).

The IKEA and radiology cases fit the distinction the Growth OS drew in 2025: "Efficiency asks what can be automated. Growth asks what can be expanded" (Puglisi, 2025e). Labor economics draws the same line. Acemoglu and Restrepo (2019) describe automation as displacing labor from tasks it already performed, and the creation of new tasks in which labor has a comparative advantage as reinstating it. At IKEA the automated task was answering routine questions, and the expanded work was design and sales. In Huang's account of radiology, the automated task is reading images and the expanded work is diagnosis, an account White disputes. Those economists add a caution. Their estimates found that across the three decades before their study, displacement accelerated and reinstatement weakened, with "rapid automation that is not being counterbalanced by the creation of new tasks" (Acemoglu and Restrepo, 2019). Neither case was run under CBG, and neither counts as a test of the framework. Both point where the Growth OS points, with AI taking the task and people growing the work the task served.

The job that grows

Read through the evidence in Part 4, the future of work in this framework is a role more than a technology. The job that expands is the governor's job: framing the problem, assigning the platforms, reading the disagreements, deciding, and answering for the decision. CBG's fourth property holds that this role builds the person who performs it, and HEQ exists to test that claim over time. The author's own ten-month record shows the Collaborative Intelligence Quotient, one of

HEQ's four dimensions, rising from 88.4 to 93.4, a single-practitioner result stated as n of 1 and nothing more (Puglisi, 2026g). The Growth OS said in 2025 that culture is the multiplier. The 2026 version names the part of the culture that multiplies: a named human who holds authority at the checkpoint and gets better at holding it.

Part 9. Governance as the Safety Net and the Cognitive Tool

Governance, in the sense this paper uses, does two jobs at once. The first is the safety net. A named human with binding authority is the point where a consequential AI outcome can still be stopped, and the record is what lets anyone prove afterward who stopped it or let it through. The author's work sets one boundary on that authority, drawn from Isaac Asimov's Three Laws of 1942 and the Zeroth Law of 1985: no governor may direct an AI-assisted outcome that injures a person, allows harm through inaction, or harms humanity (Puglisi, 2025f). The constitution explains what the boundary does for that authority: "the governor's authority is supreme within that boundary, and the boundary is what makes that authority legitimate rather than arbitrary" (Puglisi, 2026p). *The AI HOAX Distraction* carried the same logic into the policy fight over frontier AI. "Capability determines what a system can do. Autonomy determines what it can do without asking." The implementable ask was narrow: no AI system should hold authority to act without "a named human accountable at the decision point and a record of what it did" (Puglisi, 2026h).

The second is the cognitive tool. The injection function and the developmental mechanism in Part 6 make the checkpoint the place where human knowledge enters the work and where the reviewer is built. The November 2025 article made the same point about learning: "Governance becomes the pedagogy behind AI use, in the same way that instructional design became the pedagogy behind computers in classrooms" (Puglisi, 2025j). Governance built this way asks the person to think at the moment the machine would let them stop, and it measures whether they are getting better at it.

Both jobs serve the same end. The safety net keeps people answerable for what AI does, and the cognitive tool keeps people capable of answering. Together they keep AI the amplifier and not the replacement, which is the condition the Growth OS in Part 8 depends on. In 1948 Wiener described the moral position of those contributing to the new science of cybernetics as, at the least, not very comfortable. Seventy-eight years later, that discomfort has a practical answer that does not depend on predicting what the machines will become. Someone designed the work, someone set the boundaries, and someone decides what leaves the room. Human-AI Collaboration, as this practice defines it, means that person is named, holds the authority to stop the work, and can prove afterward what was decided and why.

Sources

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives, 33*(2), 3-30. https://doi.org/10.1257/jep.33.2.3

Anderson, M. M., & Fort, K. (2022). Human where? A new scale defining human involvement in technology communities from an ethical standpoint. *International Review of Information Ethics, 31*. https://informationethics.ca/index.php/irie/article/download/477/449

Anthropic. (2026, May 25). How we contain Claude across products. https://www.anthropic.com/engineering/how-we-contain-claude

Article 29 Data Protection Working Party. (2018). *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679* (WP251rev.01; adopted October 3, 2017, as last revised and adopted February 6, 2018). https://www.aepd.es/documento/wp251rev01-en.pdf

Article 36. (2013, April 19). Killer robots: UK Government policy on fully autonomous weapons. https://article36.org/statements-updates/killer-robots-uk-government-policy-on-fully-autonomous-weapons-2

Association for Computing Machinery. (1992, October 16). *ACM code of ethics and professional conduct (1992)*. https://www.acm.org/code-of-ethics/1992-acm-code

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies, 66*(9), 688-699. https://doi.org/10.1016/j.ijhcs.2008.06.001

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

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122 (Correction published August 20, 2025, *PNAS, 122*(34), e2518204122, revising an author affiliation.)

Bloomberg, J. (2026, May 31). Why 'human in the loop' falls short – and what to do about it. *SiliconANGLE*. https://siliconangle.com/2026/05/31/human-loop-falls-short/

Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency. (2011, April 4). *Supervisory guidance on model risk management* (SR 11-7). https://www.federalreserve.gov/boarddocs/srletters/2011/sr1107a1.pdf

Boinodiris, P., & Mackenzie, J. (2026, June 17). *Why "human in the loop" alone is not a governance strategy*. IBM Think. https://www.ibm.com/think/insights/liability-laundering-problem-human-in-the-loop-not-governance-strategy

Brynjolfsson, E., Chandar, B., & Chen, R. (2026). *Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence* [Working paper, revised August 12, 2026]. Stanford

Digital Economy Lab. https://digitaleconomy.stanford.edu/app/uploads/2026/08/Canaries_August2026.pdf

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1-21. https://doi.org/10.1145/3449287

Bynum, T. (2015). Computer and information ethics. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Summer 2020 ed.). https://plato.stanford.edu/archives/sum2020/entries/ethics-computer/

Corporate Responsibility for Financial Reports, 15 U.S.C. § 7241 (2002). Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, § 302, 116 Stat. 777. https://www.law.cornell.edu/uscode/text/15/7241

Crist, C. (2025, September 22). Nearly 4 in 10 companies will replace workers with AI by 2026, survey shows. *HR Dive*. https://www.hrdive.com/news/companies-will-replace-workers-with-ai-by-2026/760729/

de Winter, J. C. F., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work, 16*(1), 1-11. https://doi.org/10.1007/s10111-011-0188-1

Doerer, K. (2025, May 9). Klarna changes its AI tune and again recruits humans for customer service. *CX Dive*. https://www.customerexperiencedive.com/news/klarna-reinvests-human-talent-customer-service-AI-chatbot/747586/

Electronic Records; Electronic Signatures, 21 C.F.R. § 11.10(e) (1997). 62 Fed. Reg. 13464 (March 20, 1997). https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society, 5*, 40-60. https://doi.org/10.17351/ests2019.260

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394. https://doi.org/10.1518/001872095779064555

Engelbart, D. C. (1962). *Augmenting human intellect: A conceptual framework*. Stanford Research Institute.

European Economic and Social Committee. (2017, May 31). *Opinion on artificial intelligence* (INT/806), OJ C 288. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016IE5369

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Article 22. https://eur-lex.europa.eu/eli/reg/2016/679/oj

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 (Artificial Intelligence Act), Articles 14 and 26. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

EY. (2026, September 15). EY survey finds that autonomous AI implementation outpaces oversight, yielding an AI governance gap. https://www.ey.com/en_us/newsroom/2026/09/ey-survey-finds-that-autonomous-ai-implementation-outpaces-oversight-yielding-an-ai-governance-gap

Failure of Corporate Officers to Certify Financial Reports, 18 U.S.C. § 1350 (2002). Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, § 906, 116 Stat. 806. https://www.law.cornell.edu/uscode/text/18/1350

France. (1978, January 6). *Loi n° 78-17 relative à l'informatique, aux fichiers et aux libertés*, Article 2 (English text of the original). https://www.ssi.ens.fr/textes/a78-17-text.html

GitHub. (2025). basilpuglisi/Checkpoint-Based-Governance, commit history. https://github.com/basilpuglisi/Checkpoint-Based-Governance/commits

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, Article 105681. https://doi.org/10.1016/j.clsr.2022.105681

Green, B., & Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), Article 50. https://doi.org/10.1145/3359152

Horowitz, M. (2025, May 22). Autonomous weapon systems: No human-in-the-loop required, and other myths dispelled. *War on the Rocks*. https://warontherocks.com/autonomous-weapon-systems-no-human-in-the-loop-required-and-other-myths-dispelled/

IBM Institute for Business Value. (2026, June 8). New IBM study finds CIOs and CTOs face growing AI control gap as enterprise deployment scales. https://newsroom.ibm.com/2026-06-08-new-ibm-study-finds-cios-and-ctos-face-growing-ai-control-gap-as-enterprise-deployment-scales

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*, 4-11. https://doi.org/10.1109/THFE2.1960.4503259

McRuer, D. T., & Krendel, E. S. (1959). The human operator as a servo system element. *Journal of the Franklin Institute, 267*(5), 381-403. https://www.sciencedirect.com/science/article/pii/0016003259900912

Moray, N., Rodriguez, D., & Clegg, B. (2000). Levels of automation in process control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 44*(1), 93-96. https://doi.org/10.1177/154193120004400125

Morgan Stanley. (2026, August 27). How boards and executives are governing the rise of AI. https://www.morganstanley.com/insights/articles/ai-governance-adoption-report-2026

National Park Service. (n.d.). Stanislav Petrov. https://www.nps.gov/people/stanislav_petrov.htm

OECD. (1980, September 23; revised 2013). *Guidelines governing the protection of privacy and transborder flows of personal data*. https://legalinstruments.oecd.org/public/doc/114/body-text.en.html

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381-410. https://doi.org/10.1177/0018720810376055

Passi, S., & Singh, R. (2026, July 29). *The oversight fallacy: Why AI agents require more than humans-in-the-loop*. Data & Society. https://datasociety.net/research-library/the-oversight-fallacy-why-ai-agents-require-more-than-humans-in-the-loop/

Puglisi, B. C. (2024, February 1). Factics make us more intelligent. basilpuglisi.com. https://basilpuglisi.com/factics-make-us-more-intelligent/

Puglisi, B. C. (2025a, August 11). Content disclaimer [Page retired; content now on the Ethics of AI Disclosure page]. basilpuglisi.com. https://basilpuglisi.com/content-disclaimer/

Current version: Puglisi, B. C. (2026, September 15). Ethics of AI disclosure. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/ai-ethics/

Puglisi, B. C. (2025b, August 29). The Growth OS: Leading with AI beyond efficiency. basilpuglisi.com. https://basilpuglisi.com/the-growth-os-leading-with-ai-beyond-efficiency/

Puglisi, B. C. (2025c, September 4). The Growth OS: Leading with AI beyond efficiency, Part 2. basilpuglisi.com. https://basilpuglisi.com/the-growth-os-leading-with-ai-beyond-efficiency-part-2/

Puglisi, B. C. (2025d, September 15). How 5 AI tools drive my content strategy [Post]. LinkedIn. https://www.linkedin.com/posts/basilpuglisi_how-5-ai-tools-drive-my-content-strategy-activity-7373497926997929984-2W8w

Puglisi, B. C. (2025e, September 18). The human advantage in AI: Factics, not fantasies. basilpuglisi.com. https://basilpuglisi.com/the-human-advantage-in-ai-factics-not-fantasies/

Puglisi, B. C. (2025f, September 23). Checkpoint-Based Governance: An implementation framework for accountable human-AI collaboration (v2 drafting). basilpuglisi.com. https://basilpuglisi.com/checkpoint-based-governance-an-implementation-framework-for-accountable-human-ai-collaboration-v2-drafting/

Current version: Puglisi, B. C. (2026, September 11; updated September 23). What is Checkpoint-Based Governance? Human authority. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/cbg/

Puglisi, B. C. (2025g, September 26). The HAIA-RECCLIN model: A comprehensive framework for human-AI collaboration (draft). basilpuglisi.com. https://basilpuglisi.com/the-haia-recclin-model-a-comprehensive-framework-for-human-ai-collaboration-draft/

Current version: Puglisi, B. C. (2026, September 11). What is HAIA-RECCLIN? Reasoning and Dispatch. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/haia-recclin/

Puglisi, B. C. (2025h, October 18). The real AI threat is not the algorithm. It's that no one answers for the decision. basilpuglisi.com. https://basilpuglisi.com/the-real-ai-threat-is-not-the-algorithm-its-that-no-one-answers-for-the-decision/

Puglisi, B. C. (2025i, November). *Governing AI: When capability exceeds control* (ISBN 9798349677687). Digital Ethos. https://basilpuglisi.com/governing-ai-when-capability-exceeds-control/

Puglisi, B. C. (2025j, November 26). The methodology problem: Why research on AI and cognition confounds technology without governance use. basilpuglisi.com. https://basilpuglisi.com/the-methodology-problem-why-research-on-ai-and-cognition-confounds-technology-without-governance-use/

Puglisi, B. C. (2025k, November 14; updated March 17, 2026). HAIA-RECCLIN v2: The multi-AI governance framework for individuals, businesses & organizations (Responsible AI Growth Edition). basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/haia-recclin-v2/

Puglisi, B. C. (2026a, February 28). When AI acts between approvals: The gap everyone sees and no one has closed. basilpuglisi.com. https://basilpuglisi.com/when-ai-acts-between-approvals-the-gap-everyone-sees-and-no-one-has-closed/

Puglisi, B. C. (2026b, May 7). The Other AI: Augmented intelligence and the honest future of human-AI collaboration. basilpuglisi.com. https://basilpuglisi.com/the-other-ai-augmented-intelligence-governance/

Puglisi, B. C. (2026c, June 4). Why agentic AI was always going to fail. basilpuglisi.com. https://basilpuglisi.com/why-agentic-ai-was-always-going-to-fail/

Puglisi, B. C. (2026d, September 2). Factics: The method behind valuable content and trusted AI. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/factics/

Puglisi, B. C. (2026e, April 28; updated September 15). HAIA: The Human Artificial Intelligence Assistant ecosystem. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/haia-the-human-artificial-intelligence-assistant-ecosystem/

Puglisi, B. C. (2026f, September 11). What is HAIA-CAIPR? Cross AI Platform Review. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/haia-caipr/

Puglisi, B. C. (2026g, September 13). What is HEQ? AI literacy assessment for governing AI. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/heq-ais/

Puglisi, B. C. (2026h, September 17). The AI HOAX distraction: Safety, money, and why you can't tell which is driving. basilpuglisi.com. https://basilpuglisi.com/ai-hoax/

Puglisi, B. C. (2026i, May 1). *The AI cognitive decline narrative has not tested what it claims: A methodological audit of AI cognition research, with HAIA-RECCLIN Reasoning and HEQ with AIS composite as testable counter-proposals and a five-arm validation design*. SSRN. https://doi.org/10.2139/ssrn.6686498

basilpuglisi.com version: Puglisi, B. C. (2026, May 4). The AI cognitive decline narrative has not tested what it claims: A methodological audit of AI cognition research, with HAIA-RECCLIN Reasoning and HEQ with AIS composite as testable counter-proposals and a five-arm validation design. basilpuglisi.com. https://basilpuglisi.com/ai-cognitive-decline-narrative-untested/

Puglisi, B. C. (2026j, May 6). How credentialed professionals shape policy when method governance is stripped [The Horvath case study]. basilpuglisi.com. https://basilpuglisi.com/how-credentialed-testimony-outpaces-research-horvath-case-study/

Puglisi, B. C. (2026k, June 1). Stop blaming AI for what the education system abandoned. basilpuglisi.com. https://basilpuglisi.com/stop-blaming-ai-for-education/

Puglisi, B. C. (2026l, June 21). The continued failure in AI literacy: AILit produced a starting point halfway through the race and called theory a framework. basilpuglisi.com. https://basilpuglisi.com/ailit-framework-rule-optional/

Puglisi, B. C. (2026m, last updated September 15). Artificial intelligence: Technology, capability, governance, and human accountability. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/ai-artificial-intelligence/

Puglisi, B. C. (2026n, June 12). Why you cannot program or prompt governance into AI. basilpuglisi.com. https://basilpuglisi.com/program-prompt-governance-ai/

Puglisi, B. C. (2026o, September 11). *CAIPR: Cross AI platform review* (Fourth Edition, Version 6). Zenodo. https://doi.org/10.5281/zenodo.22710389

Puglisi, B. C. (2026p, March 10; updated September 28). Checkpoint-Based Governance (CBG): A constitutional framework for human-AI collaboration. basilpuglisi.com. Retrieved September 28, 2026, from https://basilpuglisi.com/checkpoint-based-governance/

Puglisi, B. C. (2026q, June 28). The on-ramp problem: What the Canaries dashboard shows, and what it cannot measure. basilpuglisi.com. https://basilpuglisi.com/ai-jobs-on-ramp-measurement/

Quiroz-Gutierrez, M. (2026, July 19). A decade after the 'Godfather of AI' said radiologists were obsolete, their salaries are up to $571K and demand is growing fast. *Fortune*. https://fortune.com/article/ai-godfather-radiologists-obsolete-salaries-up-to-571k-demand-growing/ (A version first appeared on Fortune.com on May 4, 2026.)

Responsibility and Authority of the Pilot in Command, 14 C.F.R. § 91.3(a) (current as of September 28, 2026). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-A/section-91.3

Rogan, J. (Host). (2025, December 3). #2422 - Jensen Huang [Audio podcast episode]. In *The Joe Rogan Experience*. Spotify. https://open.spotify.com/episode/0yT4ec9M6GobLC5ByN8pX3

Samuel, A. L. (1960). Some moral and technical consequences of automation: A refutation. *Science, 132*(3429), 741-742. https://doi.org/10.1126/science.132.3429.741

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI, 5*, Article 15. https://doi.org/10.3389/frobt.2018.00015

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MIT Man-Machine Systems Laboratory. DTIC ADA057655.

Singh, R., & Passi, S. (2026, August 12). *A sociotechnical research agenda for the oversight of AI agents*. Data & Society. https://datasociety.net/research-library/a-sociotechnical-research-agenda-for-the-oversight-of-ai-agents/

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*(4), 701-717. https://doi.org/10.1006/ijhc.1999.0349

Smith, R. M. (1967, July 1). *Analysis and design of space vehicle flight control systems, Volume X: Man in the loop*. NASA. https://ntrs.nasa.gov/citations/19670020803

U.S. Air Force. (2009, May 18). *Unmanned aircraft systems flight plan 2009-2047*. https://nps.edu/documents/106607930/106914584/UAS%2BAir%2BForce%2BRoadmap%2B2009.pdf/179e02eb-a788-4a85-a791-c13c39640ac9

U.S. Department of Defense, Nuclear Matters. (n.d.). Chronology. https://www.acq.osd.mil/ncbdp/nm/chronology/index.html

U.S. Department of Defense. (1985, December 26). *Trusted computer system evaluation criteria* (DoD 5200.28-STD). https://csrc.nist.gov/publications/history/dod85.pdf

U.S. Government Accountability Office. (1977, September 13). *Automated decisionmaking and computer-related crimes: A discussion of two GAO reports*. https://www.gao.gov/products/103077

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. https://doi.org/10.1038/s41562-024-02024-1

van de Sande, D., Economou-Zavlanos, N., & van Genderen, M. E. (2026). Meaningful oversight of medical AI beyond human in the loop. *npj Digital Medicine, 9*, Article 569. https://doi.org/10.1038/s41746-026-02971-1

White, B. (2025, December 3). Radiology isn't an example of Jevons paradox. benwhite.com. https://www.benwhite.com/radiology/radiology-isnt-an-example-of-jevons-paradox-yet/

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. The Technology Press; John Wiley & Sons; Hermann et Cie. Reissued 2019, MIT Press. https://direct.mit.edu/books/oa-monograph/4581/Cybernetics-or-Control-and-Communication-in-the

Wiener, N. (1950). *The human use of human beings: Cybernetics and society*. Houghton Mifflin.

Wiener, N. (1960). Some moral and technical consequences of automation. *Science, 131*(3410), 1355-1358. https://doi.org/10.1126/science.131.3410.1355

Zillman, C. (2026, July 30). Inside Ikea's big bet on humans in the age of AI. *Fortune*. https://fortune.com/2026/07/30/ikea-ai-workforce-reskilling-jobs-billie-chatbot-global-500/

Appendix. Source Provenance for Author Publications

Where this paper cites one of the author's own publications, the list below gives the outside sources that publication relies on, so a reader can trace the evidence beneath it. The author's citations to his own work are omitted from these lists. The main Sources list above holds the works this paper cites directly.

Puglisi (2025b). The Growth OS: Leading with AI beyond efficiency.

Sources: Canady, 2021; Deloitte, 2017; EY, 2024; F7i.AI, 2025; Forbes, 2025; IMEC, 2025; Innovapptive, 2025; McKinsey & Company, 2025; PwC, 2025; Stanford HAI, 2024

Puglisi (2025c). The Growth OS: Leading with AI beyond efficiency, Part 2.

Sources: Canady, 2021; Deloitte, 2017; EY, 2024; F7i.AI, 2025; Forbes, 2025; IMEC, 2025; Innovapptive, 2025; McKinsey & Company, 2025; PwC, 2025; Stanford HAI, 2024

Puglisi (2025e). The human advantage in AI: Factics, not fantasies.

Sources: Bauer et al., 2024; Brynjolfsson & Mitchell, 2017; Funke et al., 2024; Jobin et al., 2019; McKinsey & Company, 2025; NIST, 2023; Ouali et al., 2024; Rao & Bourne, 2025; Sadiq et al., 2021; van der Aalst et al., 2024; Wilson & Daugherty, 2018; Zhao et al., 2022

Puglisi (2025f). Checkpoint-Based Governance: An implementation framework for accountable human-AI collaboration (v2 drafting).

Sources: Bradley, 2025; Congruity 360, 2025; European Parliament and Council, 2024; International Telecommunication Union, 2025; ISACA, 2025; Lu et al., 2019; Lumenalta, 2025; McKinsey & Company, 2025; National Institute of Standards and Technology, 2023; Nexastack, 2025; OECD, 2025; Parasuraman & Manzey, 2010; Precisely, 2025; Splunk, 2025; Stanford Institute for Human-Centered Artificial Intelligence, 2024; Strobes Security, 2025; Superblocks, 2025; Visa, 2025

Puglisi (2025g). The HAIA-RECCLIN model: A comprehensive framework for human-AI collaboration (draft).

Sources: Ali et al., 2025; Angwin et al., 2016; Anthropic, 2025; Ashman & Sridharan, 2025; Bender et al., 2021; Brown et al., 2020; CNBC, 2025; Damschroder et al., 2009; Dastin, 2018; European Union, 2024; Glasgow et al., 1999; IEEE, 2019; IONI AI, 2025; Lamanna, 2025; Li et al., 2023; Microsoft, 2025; MIT, 2023; Mäntymäki et al., 2025; PwC, 2025; Reeves & Nass, 1996; Reinecke & Gajos, 2024; Reuters, 2025; Ross & Swetlitz, 2018; Salesforce, 2025; Taleb, 2012; The Verge, 2025; UNESCO, 2021; United Nations Secretary-General, 2025; Weiser, 2023; Windows Central, 2025; World Economic Forum, 2025; Zhang & Li, 2025

Puglisi (2025h). The real AI threat is not the algorithm. It's that no one answers for the decision.

Sources: Algorithmic Accountability Act of 2025, 2025; Angwin et al., 2016; Approveit, 2025; Clarke, 2025; Directors & Boards, 2025; European Commission, 2024; Hill, 2020; Informs Institute, 2025; McKinsey & Company, 2025; National Institute of Standards and Technology, 2019; National Institute of Standards and Technology, 2023; UNESCO, 2024

Puglisi (2025i). *Governing AI: When capability exceeds control* (ISBN 9798349677687).

Sources: AI Security Institute, 2026; Altman et al., 2023; American Psychological Association, 2026; Anderljung et al., 2023; Angwin et al., 2016; Associated Press, 2024; Atari et al., 2023; Babina, 2025; Bai, 2025; BBC Radio 4, 2024; Bellamy et al., 2018; Bengio et al., 2024; Bengio et al., 2025; Bloomberg, 2025; Bondar, 2025; Bowman, 2023; California Legislature, 2025; CalMatters, 2025;

Challenger, Gray & Christmas, 2026; Chatham House, 2025; Chen & Magramo, 2024; Châtelet, 2025; City of Seattle, 2025; Coalition for Content Provenance and Authenticity (C2PA), n.d.; Cohen & Suzor, 2024; Colorado Department of Education, 2024; Council of Europe, 2024; Cynergy Bank, 2024; DataReportal, 2024; Dutch Data Protection Authority, 2024; ElevenLabs, 2025; Eloundou et al., 2023; Emerging Technology Observatory, 2025; European Commission, 2024; European Parliament, 2024; EY, 2025; Federal Trade Commission, 2023; Financial Times, 2025; Future of Life Institute, 2025; Ge & Zhao, 2024; Google DeepMind, 2025; Grace et al., 2024; Harvard Law School, 2025; Hemmer et al., 2025; Horwitz & Seetharaman, 2021; Human Rights Watch, 2025; IBM Security, 2025; International Association of Privacy Professionals, 2025; International Biosecurity and Biosafety Initiative for Science, 2024; International Committee of the Red Cross, 2025; Kim et al., 2025; Laird et al., 2025; Larson et al., 2016; Layoffs.fyi, 2025; Liang, 2023; ManpowerGroup, 2026; Marcus, 2023; Microsoft Threat Intelligence, 2024; Microsoft, 2025; Miotti & Wasil, 2023; National Institute of Standards and Technology, 2023; National Institute of Standards and Technology, 2024; National Science Advisory Board for Biosecurity, 2023; Nikkei Asia, 2026; OECD, 2025; OpenAI, 2025; OpenAI, 2026; Park et al., 2024; Pew Research Center, 2025; RAND Corporation, 2025; Reuters, 2025; Stanford HAI and RegulatingAI, 2024; Stanford Institute for Human-Centered Artificial Intelligence, 2026; Surfshark, 2026; The Stanford Institute for Human-Centered AI (HAI), 2024; The WHO Team, 2022; U.S. Bureau of Labor Statistics, 2025; U.S. Government, 2024; UK Government AI Safety Institute, 2024; United Nations General Assembly, 2024; United Nations News, 2025; United Nations Security Council, 2021; Urbina et al., 2022; US Department of Defense, 2023; Vaccaro et al., 2024; War on the Rocks, 2025; Wilson, 2024; Yudkowsky, 2022

Puglisi (2025j). The methodology problem: Why research on AI and cognition confounds technology without governance use.

Sources: Crawfurd et al., 2025; Cuban, 2001; Diaz, 2025; Flavell, 1979; Johnson & Johnson, 1999; Kahneman, 2011; Kosmyna et al., 2025; Lee et al., 2025; Liu & Wang, 2024; Schraw & Dennison, 1994; Sweller, 2010; Vygotsky, 1978

Puglisi (2025k). HAIA-RECCLIN v2: The multi-AI governance framework for individuals, businesses & organizations (Responsible AI Growth Edition).

Sources: Actian, 2025; Adepteq, 2025; Anthropic, 2023; Anthropic, 2024; Anthropic, 2025; Australian Government Department of Industry, Science and Resources, 2025; Bito.ai, 2024; Bloomberg, 2025; Business Standard, 2025; Center for AI Safety, 2023; CFO Tech Asia, 2023; Cloud Revolution, 2025; Cloud Wars, 2024; CNBC, 2023; CNBC, 2025; Cooper et al., 2025; CRN, 2025; Data Studios, 2025; Deloitte, 2025; Deloitte AI Institute, 2024; Dr. Ware & Associates, 2024; Entrepreneur, 2023; European Data Protection Supervisor, 2025; EY, 2024; EY, 2025; Forrester Research, 2024; Fortune, 2025; Galileo AI, 2025; Gartner, 2025; GeekWire, 2025; Governance Institute of Australia, 2024; Hinton, 2023; Hinton, 2024; IDC, 2024; IT Channel Oxygen, 2024; Latent Space, 2024; Leone, 2025; Lighthouse Global, 2025; LinkedIn, 2024; LinkedIn, 2025; Meet Cody AI, 2023; Metomic, 2025; Microsoft, 2024; Microsoft, 2025; Microsoft Corporation, 2025; Mobile World Live, 2024; OpenAI, 2025; Parokkil et al., 2024; Partner Microsoft, 2024; PwC, 2025; Radiant Institute, 2024; Rao et al., 2025; Reddit, 2023; Reuters, 2025; Riva, 2025; Schwartz et al., 2022; SiliconANGLE, 2025; Spataro, 2025; TechCrunch, 2025; Technology Record, 2024; UC Today, 2024; UNESCO, 2024; Wall Street Journal, 2024; Yahoo Finance, 2025

Puglisi (2026a). When AI acts between approvals: The gap everyone sees and no one has closed.

Sources: Cyber Security Agency of Singapore, 2025; European Union, 2024; Madkour et al., 2026; National Institute of Standards and Technology, 2023; Vellum AI, 2025; World Economic Forum, 2024

Puglisi (2026b). The Other AI: Augmented intelligence and the honest future of human-AI collaboration.

Sources: Anthropic, 2025; Atari et al., 2023; Bartlett, 2025; Bartlett, 2026; Bastani et al., 2025; Bender et al., 2021; Brynjolfsson et al., 2025; Chalk and Talk Podcast, 2026; Dellermann et al., 2019; Dweck, 2006; Engelbart, 1962; EY, 2025; Hao, 2025; Harvard Crimson, 2023; Hemmer et al., 2025; Kahneman, 2011; Kasparov, 2017; Lane et al., 2025; Licklider, 1960; Marcus & Davis, 2019; McKinsey & Company, 2025; Noy & Zhang, 2023; PwC, 2025; Resume.org, 2025; Tetlock & Gardner, 2015; Vaccaro et al., 2024; World Economic Forum, 2025

Puglisi (2026c). Why agentic AI was always going to fail.

Sources: 14 CFR 91.3; 21 CFR 11.100; Anthropic Institute, 2026; Blum, 2010; Carnegie Mellon University, 2025; Cloudflare / Prince, 2026; Colorado General Assembly, 2024; Connecticut General Assembly, 2026; Council of Europe, 2024; Economist/YouGov, 2026; EDPB/WP29, 2018; Electronic Privacy Information Center, 1999; European Commission, Council, Parliament, 2026; European Union, 2016; European Union, 2024; EY, 2025; Financial Crisis Inquiry Commission, 2011; Gartner, 2025; Google, 2025; Grant Thornton, 2026; Hao, 2025; Harris, 2025; Harvard Business School, 2025; Human Statement / Future of Life Institute, 2026; IBM, 2026; IEEE Spectrum, 2026; ISO/IEC, 2023; ITIF/Morning Consult, 2026; Microsoft, n.d.; MIT NANDA, 2025; NIST, 2023; OECD, 2019; OpenClaw, 2026; PCAOB, 2024; Pew Research Center, 2023; RAND, 2024; S&P Global Market Intelligence, 2025; Saeri et al., 2026; Salesforce, 2025; Sarbanes-Oxley Act; The Moltbook Illusion, 2026; UK ICO, n.d.; United Kingdom, 2025; xAI v. Weiser, 2026

Puglisi (2026h). The AI HOAX distraction: Safety, money, and why you can't tell which is driving.

Sources: Amodei, 2026; Anthropic PBC v. Department of Defense, 2026; Boston Consulting Group, 2025; Directive (EU) 2024/2853; Future of Life Institute, 2026; Gartner, 2025; Kim et al., 2025; Kohli, 2026; Leung et al., 2026; Lior, 2025; McKinsey & Company, 2025; METR & Redwood Research, 2026; MIT NANDA Initiative, 2025; Olson, 2026; OpenAI, 2019; OpenAI, 2026; Parasuraman & Riley, 1997; Reuters, 2026; Trump, 2026; W. R. Berkley Corporation, n.d.; Zuckerberg, 2026

Puglisi (2026i). The AI cognitive decline narrative has not tested what it claims: A methodological audit of AI cognition research, with HAIA-RECCLIN Reasoning and HEQ with AIS composite as testable counter-proposals and a five-arm validation design.

Sources: Anderson & Krathwohl, 2001; Bastani et al., 2025; Bloom et al., 1956; Buçinca et al., 2021; Chi & Wylie, 2014; Dellermann et al., 2019; Doshi & Hauser, 2024; Dweck, 2006; Engelbart, 1962; Festinger, 1957; Flavell, 1979; Freeman et al., 2014; Fütterer et al., 2026; Ganuthula & Balaraman, 2025; Gardner, 1983; Garg et al., 2025; Gerlich, 2025; Hoffmann et al., 2014; Hopewell et al., 2025; Krathwohl, 2002; Larsen et al., 2022; Lee & See, 2004; Lee et al., 2025; Licklider, 1960; Liu et al., 2020; Noy & Zhang, 2023; Risko & Gilbert, 2016; Sidra & Mason, 2026; Skivington et al., 2021; Sparrow et al., 2011; Sterne et al., 2019; Thurn et al., 2023; Vaccaro et al., 2024; Vaidis & Bran, 2019; Vasconcelos et al., 2023; Vered et al., 2023; Vygotsky, 1978; Wekerle et al., 2024; Zhai et al., 2024

Puglisi (2026j). How credentialed professionals shape policy when method governance is stripped.

Sources: Associated Press, 2024; Atari et al., 2023; Bratsberg & Rogeberg, 2018; C-SPAN, 2026; Carr, 2010; Chalkbeat, 2026; Committee on Publication Ethics, n.d.; Goodreads, 2025; Hattie, 2023; Henrich et al., 2010; Hinton, 2023; Horvath, 2026; International Committee of Medical Journal Editors, 2026; Jiang & Loewen, 2021; Jiang et al., 2024; Juliani, 2026; KCTV5 News, 2026; Khalifeh et al., 2026; Missouri House of Representatives, 2026; OECD, 2023; Pajak, n.d.; Smith et al., 2024; Society of Professional Journalists, 2014; Stokke & Horvath, 2026; U.S. Senate Commerce Committee, 2026; Wei & Zhang, 2025; Ziernwald et al., 2022

Puglisi (2026k). Stop blaming AI for what the education system abandoned.

Sources: Alvero et al., 2024; Dizon et al., 2026; Doshi & Hauser, 2024; Fan et al., 2025; Hattie, 2009; Horvath, 2026; Moon et al., 2025; Moon et al., 2026; U.S. Senate Committee on Commerce, Science, and Transportation, 2026; Winthrop, 2026

Puglisi (2026l). The continued failure in AI literacy: AILit produced a starting point halfway through the race and called theory a framework.

Sources: Ansari, 2026; Article 29 Working Party, 2018; Atari et al., 2023; Bastani et al., 2025; Batool et al., 2025; Chi & Wylie, 2014; Court of Justice of the European Union, 2023; Eurostat, 2026; Fan et al., 2025; Festinger, 1957; Flavell, 1979; Ganuthula & Balaraman, 2025; Gerlich, 2025; Henrich et al., 2010; Jobin et al., 2019; Mullin, 2026; OECD & European Commission, 2026; OECD Education, 2026; OECD, 2025; OECD, 2026; Regulation (EU) 2016/679; Regulation (EU) 2024/1689; Risko & Gilbert, 2016; Sidra & Mason, 2025; Topaz et al., 2026; U.S. Department of Labor, 2026; Vaccaro et al., 2024; Vygotsky, 1978; Wood et al., 1976; Xu et al., 2026

Puglisi (2026m). Artificial intelligence: Technology, capability, governance, and human accountability.

Sources: European Union, 2024; National Institute of Standards and Technology, 2023; OECD, 2024; Stanford Institute for Human-Centered Artificial Intelligence, 2026

Puglisi (2026n). Why you cannot program or prompt governance into AI.

Sources: Anthropic, 2026; Colorado General Assembly, 2026; Data (Use and Access) Act 2025; European Union, 2024; Information Commissioner's Office, n.d.; Madkour et al., 2026; OpenAI, 2026; UK Government, n.d.

Puglisi (2026o). *CAIPR: Cross AI platform review* (Fourth Edition, Version 6).

Sources: Basile et al., 2021; Cavalcante Siebert et al., 2023; Davani et al., 2022; Engin, 2025; European Union, 2024; European Union, 2026; International Organization for Standardization, 2023; International Organization for Standardization et al., 2022; Janssen, 2026; Jiang et al., 2025; Kim et al., 2025; Kohli, 2026; Linux Foundation, 2024; National Institute of Standards and Technology, 2023; Perplexity, 2026; Santoni de Sio & van den Hoven, 2018; Shu, 2026; Spiro, 2026; United States Copyright Office, 2025; Verga et al., 2024; Wataoka et al., 2024

Puglisi (2026q). The on-ramp problem: What the Canaries dashboard shows, and what it cannot measure.

Sources: Brynjolfsson, 2022; Brynjolfsson et al., 2021; Brynjolfsson et al., 2025; Lichtenberg, 2026; Stanford Digital Economy Lab, 2026a; Stanford Digital Economy Lab, 2026b

Disclaimer

The author is not a lawyer, and this paper does not provide legal advice. This is thought research and governance analysis based on public sources, cited materials, and human-AI review. It is intended to help executives, practitioners, insurers, and governance teams think more clearly about AI risk, liability exposure, and documentation practices. Readers should not rely on this paper as a legal opinion, compliance determination, or substitute for qualified counsel. Any organization facing a legal, regulatory, contractual, or insurance question should consult its own attorney, broker, or professional adviser before acting.

The regulatory, incident, and survey material in this paper is current as of September 2026 and should be re-verified before reuse. The author is an independent practitioner and author who may profit in other ways from research and content like this.

The Other AI: Audio Briefings on Augmented Intelligence and AI Governance

Spotify | Apple Podcasts | Amazon Music | YouTube Playlist