Monkeys with Typewriters: 2025 America's First Draft of a People–AI Civilization

Douglas C. Youvan

doug@youvan.com

www.youvan.ai

December 3, 2025

The year 2025 in the United States will not be remembered as the moment AI took over, nor as the moment it was decisively tamed. It will look more like the year we gave monkeys typewriters. Powerful generative models were handed to tens of millions of people who did not understand how they worked, what they were good for, or how they might quietly rewire work, politics, and interior life. Text, images, and code poured into the world at near-zero marginal cost, while the filtering, editing, and apprenticeship structures that once turned raw expression into durable knowledge lagged far behind. This paper treats "monkeys with typewriters" as a structural description rather than a sneer. It proposes a three-stage developmental framework for people–AI relations: Stage 1 monkeys (untrained mass use), Stage 2 apprentices in a scriptorium (emergent craft and norms), and Stage 3 centaurs (human–AI dyads intentionally aimed at living truth, τ(t), rather than mere engagement). Using this lens, we map what could have happened in 2025 and what in fact did, across deployment, governance, labor, epistemic order, and culture. The goal is not nostalgia for pre-AI life, but design: how to move from stochastic noise toward an intentional people–AI civilization.

Keywords: AI, artificial intelligence, 2025, United States, monkeys with typewriters, information overload, information ecosystem, logostics, centaurs, apprenticeship, human AI collaboration, epistemic environment, labor, governance, culture, ethics, τ(t), public sphere, rage bait.

A collaboration with GPT-5.1-Thinking-Extended. 57 pages. CC4.0.

Outline

1. Introduction: The Year of the Monkey-Typewriter

1.1 The metaphor of monkeys with typewriters
1.2 First contact with everyday AI, not superintelligence
1.3 From technical capability to social reality: what this paper attempts
1.4 Overview of the three-stage framework: monkeys, apprentices, centaurs

2. The 2025 Possibility Box: What Could Have Happened

2.1 Axes of possibility: deployment, governance, labor, epistemics, culture
2.2 High-regulation, low-adoption worlds that did not materialize
2.3 Runaway-adoption, crisis worlds that also did not appear (yet)
2.4 Locating the actually existing 2025 inside this possibility box

3. Stage 1: The Monkey Regime as Structural Condition

3.1 Near-zero marginal cost for text, images, and code
3.2 Monkeys, typewriters, and missing theories of use
3.3 The illusion of competence: activity mistaken for authorship
3.4 Engagement as editor: why virality, not truth, selected outputs

4. Micro-Level Monkeys: How Ordinary People Used AI in 2025

4.1 Prompt mashing as a new folk practice of writing
4.2 Homework, emails, sermons, and spam: mundane use-cases
4.3 Cognitive outsourcing and the erosion of felt authorship
4.4 Anxiety, delight, and the early phenomenology of AI dependence

5. Missing Editors: Institutions as Absent or Confused Gatekeepers

5.1 Newsrooms, publishers, and universities under AI pressure
5.2 Platforms as de facto editors: recommendation systems and rage bait
5.3 Professional guilds, ethics codes, and their slow response
5.4 Why no one really owned the problem of filtering AI-saturated content

6. The Meso Layer: Organizations as Monkey Farms

6.1 Corporate enthusiasm and superficial “AI transformation”
6.2 AI in bureaucracies: more forms, more reports, not more wisdom
6.3 Schools and churches: from bans to quiet adoption
6.4 Case sketches of emergent people–AI practices in specific sectors

7. Proto-Apprenticeships: Where Craft Quietly Emerged

7.1 The rise of informal prompt craftsmanship and workflow design
7.2 Human–AI co-writing in research, law, and creative work

1. Introduction: The Year of the Monkey-Typewriter

The year 2025 in the United States did not look like the familiar science fiction scenarios in which a single superintelligent system awakens, seizes control, or benevolently steers humanity. Instead, it looked like something far less cinematic and far more chaotic. Millions of people acquired access to powerful generative systems that could produce text, images, code, and audio at the cost of a few keystrokes. The result was not a sudden flowering of collective wisdom, nor an immediate collapse into automated madness, but a noisy, uneven, and often comical first draft of a people–AI civilization. The metaphor of giving monkeys typewriters captures this condition: capability without apprenticeship, output without clear criteria of value, and an information environment saturated with artifacts that few actors felt responsible for curating.

This paper argues that this early phase is not an accident at the margins of history but a structural stage in the co-evolution of people and AI. It proposes a simple three-stage framework to make sense of 2025 as it was lived and as it might have been otherwise: Stage 1, the monkey regime, in which access dramatically outpaces craft; Stage 2, the scriptorium of apprentices, in which norms and training begin to catch up; and Stage 3, the centaur horizon, in which human–AI dyads become intentional, accountable agents aimed at more than engagement or convenience. The purpose is not to nostalgically condemn 2025 as a fallen state, but to treat it as a first draft that can be revised if we are willing to name its patterns and design exits from them.

1.1 The metaphor of monkeys with typewriters

The image of monkeys with typewriters is usually invoked to illustrate the power of randomness: given infinite time, randomly striking keys could in principle produce Shakespeare. In this paper, the metaphor is repurposed as a description of how powerful generative tools entered everyday life in 2025. Monkeys stand for untrained users, institutions, and even the models themselves, acting without a shared theory of what they are doing. The typewriters stand for generative interfaces that accept almost any prompt and deliver fluent output regardless of whether the user has a clear purpose, adequate expertise, or a sense of responsibility for downstream consequences.

Three aspects of the metaphor are especially important. First, the cost of producing long, fluent artifacts collapses toward zero, so there is very little natural friction limiting the volume of output. Second, the selection mechanism is external to the monkeys and the typewriters; it lives in the attention environment, where relatively crude metrics such as clicks, shares, and time-on-page determine which outputs are amplified. Third, the monkeys do not have a robust sense of authorship or craft. They are rewarded for speed, novelty, and quantity, not for truthfulness,

coherence, or long-term utility. Under these conditions, it is unsurprising that the public sphere in 2025 felt saturated with text and images that were technically impressive yet epistemically fragile.

The metaphor is not meant as a sneer at individuals. Many of the monkeys are highly educated professionals trying to cope with new tools under time pressure and institutional ambiguity. Rather, the image points to an arrangement in which the dominant incentives and infrastructures make monkey-like behavior the path of least resistance. When systems are built so that anyone can generate anything without much guidance or accountability, it is structurally predictable that monkey-typewriter dynamics will dominate.

1.2 First contact with everyday AI, not superintelligence

Much of the pre-2025 public conversation about artificial intelligence revolved around extremes: either narrow tools that automate specific tasks, or hypothetical superintelligences that might surpass human capabilities in all domains. The reality of 2025 in the United States falls in between. What emerged was a broad layer of everyday AI, embedded into search engines, office suites, messaging platforms, and phones. These systems were not omniscient minds; they were large-scale text and media generators trained on massive datasets, capable of plausible continuation and synthesis but lacking a unitary worldview, long-term memory of particular users, or independent goals.

This first contact with everyday AI is significant for several reasons. It reconfigured who had access to powerful generative capabilities, extending them far beyond specialist communities. It changed the cost structure of many intellectual and creative tasks, making it simpler to draft, translate, summarize, and rephrase. It also blurred distinctions between human and machine authorship, since many artifacts were produced through rapid back-and-forth between a person and a model, with no clear boundary between where one ended and the other began. Yet despite these changes, the underlying social and institutional frameworks remained largely oriented toward a pre-AI world, in which authorship, expertise, and responsibility mapped more directly onto individual humans.

The central claim here is that the novelty of 2025 lay less in the raw technical capability of the models, and more in the sudden, mass-scale intertwining of these capabilities with everyday life. The distance between what the systems could do in principle and what societies knew how to do with them in practice defines the gap this paper explores.

1.3 From technical capability to social reality: what this paper attempts

Technical accounts of generative models typically focus on architectures, training regimes, benchmarks, and performance metrics. Social commentary often oscillates between alarm and enthusiasm. This paper does something different. It treats 2025 as a contingent social reality

shaped by interplay among technical possibility, institutional inertia, economic incentives, and cultural narratives. The question is not simply what the models could do, but what people, organizations, and infrastructures actually did with them, and what they failed to do.

The analysis proceeds at three levels. At the micro level, it examines how individual people interacted with AI systems, how they experienced authorship, and how they negotiated trust and dependence. At the meso level, it looks at organizations and professions, including firms, schools, churches, and media outlets, and asks how they incorporated or resisted AI-generated artifacts. At the macro level, it considers governance, regulation, and cultural framings that set the broad conditions for people–AI interaction. Across these levels, it uses the monkeys-with-typewriters metaphor to identify patterns of unstructured experimentation and weak filtering, and it contrasts these with counterfactual trajectories in which apprenticeship, craft, and explicit norms might have developed more quickly.

The paper does not attempt a comprehensive empirical history of 2025. Instead, it proposes a conceptual framework designed to be precise enough to guide future research and flexible enough to accommodate diverse cases. By mapping the distance between technical capability and social reality, it aims to clarify what is at stake in moving beyond the monkey regime.

1.4 Overview of the three-stage framework: monkeys, apprentices, centaurs

To organize the analysis, the paper uses a three-stage developmental framework for people–AI relations. The stages are not strictly chronological, nor are they mutually exclusive. Rather, they describe different regimes that can coexist at different scales and in different domains, with one or another tending to dominate.

Stage 1 is the monkey regime. In this stage, access to generative tools is widespread, costs of production are low, and norms of use are weak. Most actors treat AI as an opaque but convenient instrument. Outputs flood the information environment, and selection is largely driven by engagement and convenience. The defining features are volume, speed, and weak responsibility.

Stage 2 is the scriptorium of apprentices. In this stage, craft and discipline begin to emerge. People and institutions develop shared practices for prompting, verifying, editing, and integrating AI outputs. There is an expectation of training rather than mere access. Editors, in the broad sense of those who filter and shape information, regain prominence. Artifacts are fewer but carry higher investment of attention and care.

Stage 3 is the centaur horizon. Here, the unit of analysis is not the human or the model alone, but the human–AI dyad as an integrated agent. These centaurs are intentionally oriented toward norms and trajectories that transcend short-term engagement or economic gain. They aim at forms of truth, coherence, and ethical alignment that can be tracked over time. Rather

than being accidental by-products of convenience, centaur pairs become explicit sites of responsibility and design.

The remainder of the paper uses this framework to interpret 2025 in the United States as a predominantly stage-1 moment, with pockets of emerging stage-2 practice and early examples of stage-3 centaurs. It then draws out design principles and research questions for those who wish to accelerate the transition from a monkey-typewriter world to one in which apprenticeship and centaur agency are the norm rather than the exception.

2. The 2025 Possibility Box: What Could Have Happened

If 2025 in the United States felt chaotic, that is partly because there was no shared frame for what kind of year it might have been. The arrival of everyday AI had been anticipated in technical circles, but the contours of social response were not mapped in advance. One way to recover a sense of structure is to imagine 2025 as a point inside a "possibility box," defined by a set of axes along which different futures could have unfolded. The box is not a prediction device but a retrospective tool: it helps distinguish what was historically contingent from what was structurally likely, and it clarifies how different institutional choices, economic pressures, and cultural narratives might have pushed the year toward other corners of the space.

This section introduces the five axes that define the possibility box and explores two contrasting classes of worlds that did not materialize in 2025: high-regulation, low-adoption scenarios and runaway-adoption, crisis scenarios. It then sketches how the actually existing 2025 can be located within the box as a mixed, unstable configuration: more AI in daily life than many expected, less disruption than some feared, and far less intentionality than would be required to move beyond the monkey regime.

2.1 Axes of possibility: deployment, governance, labor, epistemics, culture

The possibility box is defined by five broad axes. These are not the only dimensions along which 2025 could be analyzed, but they capture key levers that shaped people–AI relations in practice.

The first axis is deployment. This refers to how widely and deeply generative systems are embedded in everyday life and work. At one extreme lies minimal deployment: AI remains a niche tool used by specialists in certain industries. At the other extreme lies pervasive deployment: AI is integrated into most devices, workflows, and services, such that opting out becomes difficult or costly. Deployment matters because it determines how many people become monkeys with typewriters, how many organizations must improvise new norms, and how much of the information environment is produced or filtered by models.

The second axis is governance. This covers laws, regulations, standards, and informal norms that constrain or steer AI development and use. One extreme is heavy, centralized governance, with stringent licensing, capability caps, and strong enforcement mechanisms. The other extreme is laissez-faire minimalism, with few formal constraints beyond existing law, leaving the direction of development largely to market forces and corporate decisions. Governance shapes not only what is allowed but what feels legitimate or inevitable.

The third axis is labor. This concerns how AI interacts with employment, wages, and class structure. On one side lies a world in which AI is slow to affect most jobs, remains a complement rather than a substitute, and leaves existing patterns of inequality largely intact. On the other side lies rapid automation of cognitive tasks, widespread job displacement, and a scrambling of existing class positions. Different points on this axis can generate very different social responses, from quiet adjustment to overt conflict.

The fourth axis is epistemics: the structure of knowledge, trust, and information quality. At one extreme, AI is successfully integrated into fact-checking, summarization, and research in ways that improve clarity and reduce the impact of misinformation. At the other extreme, AI amplifies confusion: deepfakes proliferate, synthetic text floods public discourse, and trust in media, science, and institutions erodes further. This axis is where the monkeys-with-typewriters metaphor bites most directly, as it concerns the selection mechanisms that decide which artifacts persist and shape beliefs.

The fifth axis is culture, particularly the cultural and spiritual narratives through which people interpret AI. On one end, AI is framed as a mundane tool, akin to a more powerful calculator or word processor. On the other, it is mythologized as savior, demon, oracle, or harbinger of the end times. Between these poles lies a range of attitudes: curiosity, anxiety, boredom, reverence, and satire. Cultural narratives determine whether AI is primarily seen as infrastructure, idol, threat, or partner, and thus influence both individual behavior and institutional choices.

By positioning 2025 along each of these axes, it becomes possible to speak more concretely about what kind of year it was, and how it differed from otherwise plausible alternatives.

2.2 High-regulation, low-adoption worlds that did not materialize

One family of counterfactual worlds in the 2025 possibility box can be described as high-regulation, low-adoption scenarios. In these worlds, strong institutional interventions, public skepticism, or slow technical progress prevent everyday AI from becoming a major factor in daily life.

In an extreme version of such a world, early incidents involving misuse, safety breaches, or political disruption trigger a sharp regulatory backlash. Legislatures enact broad moratoria on certain classes of models, strict licensing regimes for providers, and severe penalties for

unauthorized deployment. Institutions such as schools, hospitals, and public agencies are prohibited from using generative systems except under tightly controlled conditions. Consumer-facing interfaces remain limited, and many people's only experience of AI is through background systems such as spam filters or recommendation engines.

Even in a less dramatic version, cautious regulators and risk-averse corporations move slowly. Liability concerns make firms wary of exposing customers to generative tools. Professional bodies discourage their use in law, medicine, and education until clear standards are established. Adoption is largely confined to pilot projects, internal tools, and a few enthusiast communities. Public discourse continues to discuss AI as a future issue rather than a present reality.

In these high-regulation, low-adoption worlds, the monkey regime is muted. There are fewer monkeys with typewriters simply because fewer typewriters are accessible. Low deployment reduces the volume of synthetic artifacts. Editors, in the broad sense, retain more control over public communication channels, because content creation remains relatively costly and constrained. The epistemic environment does not escape its pre-AI pathologies, but it is not yet saturated with model-generated text and imagery.

These worlds did not materialize in the United States in 2025. Regulatory action was real but fragmented and often reactive. Institutions did not impose sweeping bans on AI, and market incentives pushed in favor of rapid rollout. Everyday AI arrived faster and more widely than a high-regulation, low-adoption scenario would allow.

2.3 Runaway-adoption, crisis worlds that also did not appear (yet)

At the opposite corner of the possibility box lie runaway-adoption, crisis scenarios. In these worlds, generative AI not only becomes ubiquitous but also precipitates visible breakdowns in economic stability, political order, or epistemic coherence.

One variant is an economic crisis scenario. Here, rapid adoption of AI tools leads to large-scale displacement of workers in a compressed timeframe. Entire categories of white-collar jobs erode. Institutions are unprepared to absorb the shock, and social safety nets prove inadequate. Political movements emerge around resistance to AI, and governments respond with emergency measures that may include restrictions, subsidies, or nationalizations. AI is no longer a noisy background factor but a central axis of conflict.

Another variant is an epistemic crisis scenario. Synthetic media and persuasive text become so sophisticated and cheap that coordinated disinformation campaigns overwhelm existing filters. Major elections are thrown into doubt as voters confront conflicting claims that they cannot independently verify. Financial markets are manipulated by AI-generated rumors and fabricated

reports. Trust in digital communication collapses, and people retreat to smaller, more insular communities where authenticity can be directly vouched for.

A third variant is a governance crisis scenario. States or non-state actors deploy AI in ways that lead to accidents, cyber incidents, or escalatory misperceptions in international relations. Automated systems misinterpret signals, or decision-makers over-rely on model outputs in high-stakes domains. A crisis unfolds that is widely perceived as an "AI event," prompting a rush to redesign or shut down systems.

These runaway-adoption crisis worlds did not fully manifest in 2025 either. Adoption was fast but not yet so disruptive as to produce mass unemployment or systemic economic failures. Disinformation and deepfakes emerged as real problems, but they did not trigger an outright collapse of trust or the breakdown of major institutions. States and corporations made missteps, but no single catastrophic incident defined the year as an AI-induced disaster.

The absence of these crises does not mean they are impossible, only that 2025 did not cross the thresholds necessary to activate them. The possibility box still contains these corners for future years.

2.4 Locating the actually existing 2025 inside this possibility box

With these counterfactual extremes in view, the actually existing 2025 in the United States can be located as an intermediate and unstable configuration. Along the deployment axis, 2025 sits closer to the high end than the low. Everyday AI has entered search, office tools, and consumer apps. A significant minority of workers use AI regularly, and many more encounter it indirectly. Yet deployment is uneven across regions, sectors, and classes; for many people, AI is still something that happens elsewhere.

On the governance axis, 2025 occupies a middle position characterized by active attention but incomplete structure. Governments, regulators, and professional bodies issue reports, guidelines, and preliminary rules. Some sectors face tighter scrutiny than others. However, there is no unified, comprehensive regime that decisively steers the trajectory. Governance remains a patchwork, and much of the practical rule-making occurs inside corporations that control the major models and platforms.

On the labor axis, the year reflects early-stage reconfiguration rather than crisis. AI is beginning to alter workflows, change the skills that are rewarded, and shift bargaining power in certain occupations. Some workers benefit from productivity gains and new opportunities, while others feel threatened by potential redundancy. Yet there is no clear, broad-based collapse in employment; instead, there is unease and speculation about what comes next.

Epistemically, 2025 lies in a zone of escalating noise and fragile trust. AI-generated content blends with human-generated material across social media, news, and entertainment. Fact-checking, moderation, and journalistic practices adapt only slowly. Many people report difficulty distinguishing authentic from synthetic artifacts, and overall confidence in information sources continues to erode. However, there remains enough continuity and redundancy in the system to prevent outright collapse. Institutions still function, even if more people eye them with suspicion.

Culturally, 2025 is marked by contested narratives about AI. Some communities embrace it as a powerful new tool, others fear it as a threat to jobs, democracy, or human uniqueness, and still others treat it as an amusing novelty. Religious and spiritual interpretations of AI exist but remain marginal compared to more secular frames such as innovation, competition, or risk. AI becomes a symbol in broader culture wars rather than a settled object of shared meaning.

Taken together, these positions place 2025 in a particular region of the possibility box: high and rising deployment, medium and unsettled governance, early but uneven labor impact, deteriorating but not yet collapsed epistemic structures, and fragmented cultural narratives. It is a world in which monkeys have already been given typewriters, but editors, apprenticeships, and centaur-like partnerships remain underdeveloped. The rest of the paper builds on this placement, arguing that recognizing where 2025 actually landed is a precondition for designing a path out of the monkey regime and away from the more dangerous corners of the box.

3. Stage 1: The Monkey Regime as Structural Condition

Stage 1, the monkey regime, is not simply a matter of individual ignorance or bad habits. It is a structural condition generated by the interaction of technical capability, economic incentives, and institutional lag. Once generative models became cheap enough to serve billions of tokens per day, and once interfaces were built that allowed anyone to produce long, fluent artifacts with minimal friction, a new equilibrium emerged. In that equilibrium, the default behavior is to generate first and think later, if at all. The monkey regime is thus not about personal failings; it is the predictable result of systems that reward speed and volume over reflection and care.

In this section, the monkey regime is unpacked along four dimensions. First, the near-zero marginal cost of producing text, images, and code plunges the information environment into oversupply. Second, the metaphor of monkeys and typewriters is used to highlight how missing theories of use leave both individuals and institutions improvising without shared standards. Third, the illusion of competence becomes widespread as fluent output is mistaken for deep understanding and authorship. Fourth, engagement metrics act as de facto editors, ensuring

that virality rather than truthfulness becomes the primary criterion for which artifacts gain prominence.

3.1 Near-zero marginal cost for text, images, and code

Before the widespread deployment of generative systems, producing long-form text, detailed images, and nontrivial code typically required substantial time, skill, and effort. Even in the era of word processors and digital cameras, someone had to sit down and type, compose, or program. These costs served as a kind of friction that limited how much content any individual or organization could realistically produce. The arrival of powerful generative models radically altered this landscape. With a few prompts and clicks, users could now obtain pages of text, galleries of images, or functional code snippets in seconds.

Economically, this is a shift in marginal cost toward zero. Once a model is trained and deployed, the cost of generating one more paragraph, picture, or function is tiny compared to the cost of human labor. The consequence is a dramatic increase in the volume of artifacts entering the public sphere. Blog posts, policy briefs, marketing copy, student essays, sermons, and social media updates can all be multiplied with little additional effort. For organizations, the temptation to automate content production is strong; for individuals, the barrier to experimentation collapses.

This abundance does not automatically translate into value. When generation is almost free, the bottleneck moves from production to attention and evaluation. Instead of asking whether there is enough content, the relevant questions become who will read, test, or critique it, and according to which standards. Yet the social and institutional structures responsible for filtering and validating information were not redesigned at the same speed. Editors, peer reviewers, teachers, and other gatekeepers suddenly find themselves facing a flood they were never scaled to handle. The result is an information environment where most artifacts receive little or no serious scrutiny, not because they are all equally good, but because there is simply too much to sift.

The near-zero marginal cost of generation thus sets the stage for monkey-like behavior. When writing another page or generating another draft feels almost costless, there is little built-in penalty for asking the model to produce more, regardless of whether there is a clear need. The constraints that once enforced selectivity are loosened, and the temptation to substitute effortless generative output for more demanding forms of thinking becomes constant.

3.2 Monkeys, typewriters, and missing theories of use

In the monkey regime, the dominant pattern is improvisation without a shared theory of use. People have access to powerful tools but lack a coherent understanding of what these tools are, what they are good for, and how they might reshape responsibilities. The monkeys-with-

typewriters metaphor captures this gap. Monkeys can press keys, but they do not know what writing is, let alone literature. They can produce strings of characters, but they lack any framework for judging the result.

Users in 2025 were of course more capable than literal monkeys, but they still faced a similar absence of shared theory. Most had never been taught how generative models work beyond vague metaphors of prediction and pattern-matching. Few knew the limits of the models' training data, the nature of their failure modes, or the ways in which biases and artifacts of the data could surface in outputs. Even fewer had explicit guidance on how to integrate AI-generated material into their work without eroding standards of evidence, originality, or accountability.

Institutions were not much better prepared. Organizations rolled out AI tools to employees with minimal training, often framing them as productivity boosters rather than as systems that might require new norms and checks. Schools issued conflicting directives about whether and how students could use AI, oscillating between permissive experimentation and blanket prohibitions. Professional bodies struggled to update codes of conduct rapidly enough, and many settled for vague warnings and aspirational guidelines.

The result was a proliferation of local, ad hoc practices. Individuals and teams developed their own micro-theories of use, ranging from cautious experimentation to uncritical delegation. Some treated models as brainstorming partners, others as ghostwriters, others as unreliable but entertaining toys. These practices sometimes converged on effective workflows, but they rarely crystallized into robust traditions or widely shared standards. The monkey regime is characterized by this patchwork of improvisations, undertaken without a common language for evaluating what is being done.

In such a context, the typewriter is used because it is there and because it produces something on command, not because there is a clear understanding of when and why it should be used. The missing theories of use are not merely academic abstractions. They are the absent scaffolding that would otherwise help people distinguish between situations where delegation is appropriate and situations where human judgment must remain central.

3.3 The illusion of competence: activity mistaken for authorship

One of the most pervasive psychological effects of the monkey regime is the illusion of competence. Generative models are designed to produce fluent, well-structured outputs that conform to familiar genres. They can draft emails in an appropriate tone, organize arguments into numbered points, mimic official or scientific styles, and surface relevant-sounding facts and references. When a user sees such an output in front of them, it is tempting to feel that a

difficult task has been completed, or at least that the user has played a central role in its completion.

This illusion has several layers. At the most basic level, it confuses linguistic fluency with understanding. Because the model can assemble sentences that resemble expert speech, users may overestimate its grasp of the subject matter. When they passively accept or lightly edit these outputs, they may in turn overestimate their own grasp, trusting that the apparent competence of the text reflects their own agency. The act of clicking, prompting, and approving can feel like authorship, even when the cognitive work has been minimal.

At a second layer, the illusion obscures the distinction between surface completion and substantive adequacy. A document that looks polished may nevertheless be wrong, misleading, or shallow. In pre-AI workflows, the effort required to produce a polished document often served as a crude proxy for the care invested in its content. Generative systems break that linkage. It becomes easy to produce something that looks finished without having done the underlying reasoning, research, or moral reflection. Activity, in the form of generating and submitting outputs, is mistaken for genuine achievement.

At a third layer, the illusion disrupts the sense of responsibility. When users feel only loosely attached to the content, they may also feel less responsible for its consequences. The fact that the model suggested the phrasing or the argument can become a psychological buffer against accountability. At the same time, institutional cultures that reward speed and volume over depth reinforce the illusion. If metrics focus on how many documents were delivered, how quickly emails were answered, or how many posts were made, then the ability to generate content rapidly using AI can be misread as a sign of high performance.

The monkey regime thus encourages a subtle but pervasive shift from authorship to approval. People become managers of outputs rather than authors of arguments. This is not inherently disastrous, but in the absence of explicit countermeasures, it erodes the habits of careful thinking that previously accompanied many acts of writing and decision-making. The typewriter produces, the monkey approves, and both move on.

3.4 Engagement as editor: why virality, not truth, selected outputs

In the monkey regime, the question of which artifacts matter is largely answered by engagement metrics. The editors that once controlled access to mass audiences have been partially replaced by algorithms that optimize for clicks, views, likes, shares, and dwell time. These algorithms do not care whether a piece of content was generated by a human, an AI, or a human–AI collaboration. They care about signals of attention. As generative tools lower the cost of producing content, the competition for attention becomes even more intense, and

engagement-driven systems become the primary selectors of which artifacts rise above the noise.

This selection mechanism interacts with the properties of generative models in specific ways. Models are good at producing content that mimics known high-engagement formats: provocative headlines, emotionally charged narratives, simplified explanations, and confident-sounding claims. Users who learn, consciously or not, which prompts yield such outputs can feed the system with prompts tuned for virality. Generative tools can be used to A/B test variations at scale, rapidly exploring which phrasing, framing, or imagery elicits the strongest reactions. The result is an arms race in which model-driven content is optimized not for accuracy or nuance, but for engagement.

Truth, in this environment, becomes at best a secondary criterion. It matters when inaccuracies are so blatant that they trigger backlash or loss of trust, but in many domains, especially those charged with emotion or ideology, engagement can be sustained even in the presence of factual distortions. The editor function of engagement thus tends to amplify content that is surprising, anger-inducing, or affirming of prior beliefs, regardless of whether it reflects careful reasoning or reliable evidence. Rage bait, sensationalism, and oversimplification find new allies in generative tools that can mass-produce such content on demand.

Institutions that might have counterbalanced this tendency struggled in 2025 to adapt their filters to the volume and speed of the monkey regime. Fact-checkers, moderators, and investigative journalists could not match the scale at which AI-assisted content was being produced and propagated. Platforms experimented with labels, down-ranking, and other interventions, but these measures were incremental and often contested. In the meantime, engagement-based algorithms continued to serve as the primary editors for much of the public sphere.

The central feature of stage 1, then, is not only the proliferation of monkeys with typewriters, but the presence of an editor whose criteria are misaligned with truth-seeking. When virality becomes the main pathway to visibility, and when generative tools are optimized to exploit that pathway, the information environment drifts away from reflective discourse and toward a constant churn of attention-grabbing artifacts. Any effort to move beyond the monkey regime must therefore address not only the behavior of individual users, but the engagement-driven systems that decide which outputs matter.

4. Micro-Level Monkeys: How Ordinary People Used AI in 2025

The monkey regime becomes visible most vividly at the micro level, where ordinary people interacted with AI systems through chat boxes, document editors, and mobile apps. This was not the realm of long-term strategy or national policy, but of late-night homework, hurried emails, Sunday sermon preparation, and side-hustle marketing copy. Across millions of small encounters, a new pattern of writing and thinking emerged: users typed a few words, received a flood of fluent output, and improvised a relationship to that output on the fly. The tools were marketed as assistants but often functioned as ghostwriters, tutors, or confessional walls. To understand 2025 as the year of the monkey typewriter, it is necessary to look closely at these everyday interactions.

This section examines four aspects of micro-level behavior. First, prompt mashing became a new folk practice of writing, in which users iteratively refined instructions rather than carefully composing drafts. Second, generative AI found its way into mundane use-cases including homework, emails, sermons, and spam. Third, cognitive outsourcing altered the experience of authorship and responsibility. Finally, ordinary users navigated a complex mix of anxiety and delight as they became increasingly dependent on tools they neither fully understood nor fully trusted.

4.1 Prompt mashing as a new folk practice of writing

For many people in 2025, the primary act of writing shifted from composing sentences to composing prompts. Instead of starting with a blank page and building an argument, story, or explanation line by line, users typed short instructions: “Write a polite email requesting an extension,” “Summarize this article in simple language,” “Give me five bullet points for a sermon on forgiveness,” “Make this sound professional.” When the first output was unsatisfying, they did not necessarily revise the underlying ideas; they revised the prompt: “More formal,” “Shorter,” “Add a joke,” “Use Bible verses,” “Make it sound like I’m an expert.”

This “prompt mashing” can be understood as a new folk practice of writing. It is folk because it developed informally, without formal instruction or explicit theory. Users learned from trial and error, from online tips, and from watching how others interacted with models. It is writing, but at one remove. The object being crafted is less the text itself than the instruction that will produce it. The craft lies in finding the right combination of constraints, keywords, and examples that elicit the desired style or content.

Prompt mashing has both empowering and distorting effects. On the empowering side, it allows people with weak writing skills to produce coherent documents, and it offers a way to explore multiple framings quickly. On the distorting side, it encourages a view of language as something to be tuned externally rather than inhabited from within. The iterative loop becomes “change

the prompt until the output feels right" rather than "think through what I actually want to say." Over time, this can shift attention away from the underlying ideas and toward surface features of presentation.

For many micro-level monkeys, the satisfaction of watching the model revise the text on demand overshadowed questions about where the words came from and what commitments they implied. The typewriter responded eagerly to every tweak, and that responsiveness became a central part of the new writing experience.

4.2 Homework, emails, sermons, and spam: mundane use-cases

The early history of everyday AI is not primarily a story of dramatic breakthroughs. It is a story of mundane use-cases quietly migrating into mixed human–machine workflows. Students used generative tools to draft essays, solve problem sets, and produce study guides. Some copied outputs directly; others used them as scaffolds, rewriting in their own words. The line between assistance and substitution was often blurred, leaving teachers unsure how to interpret polished assignments that did not match in-class performance.

Knowledge workers used AI for emails, reports, and presentations. A common pattern was to paste a rough outline or bullet list and ask the model to turn it into a well-structured memo or slide deck. Over time, some skipped the outline and began with prompts like "Write a report summarizing the attached" or "Draft a proposal for this client based on these notes." The model produced text that followed familiar corporate templates, complete with confident language and standardized phrases. The result was a proliferation of documents that looked professionally authored even when the human contribution was minimal.

Clergy and religious workers experimented with AI-generated sermons, prayers, and study materials. Tools could quickly provide outlines tied to specific themes, scriptural passages, or liturgical seasons. Some used these outputs as starting points, editing and personalizing them. Others relied more heavily on the model, especially under time pressure. This raised subtle questions about authenticity and inspiration: when a sermon moves people, who has spoken?

Meanwhile, the lowest-status corner of the information economy embraced generative tools enthusiastically. Spammers, scammers, and content farms used AI to produce endless variations on marketing emails, fraudulent messages, and low-quality web pages. Models made it easy to localize content to different languages, tweak phrasing to evade filters, and generate plausible-seeming explanations for dubious offers. The same engines that assisted students and professionals also powered a new wave of automated noise.

Taken together, these mundane use-cases show how deeply the monkey regime penetrated everyday life. It was not primarily about spectacular new experiences. It was about the quiet

substitution of generated text for human effort in domains where deadlines were tight, evaluation was coarse, and the consequences of error were often delayed.

4.3 Cognitive outsourcing and the erosion of felt authorship

As ordinary people leaned on AI for more tasks, they began to outsource not only the mechanical aspects of writing but portions of the cognitive work itself. When models summarized long articles, users sometimes skipped reading the originals. When models proposed arguments or counterarguments, users sometimes accepted them without fully understanding their structure. When models suggested ways to respond to criticism or praise, users sometimes adopted the suggested tone without asking whether it reflected their own intentions.

This cognitive outsourcing changed the experience of authorship. For some, it brought relief. They felt less overwhelmed by information, less burdened by the need to craft every sentence, and more capable of meeting the demands of work and school. For others, it introduced a quiet unease. They began to feel that the texts associated with their name were not entirely “theirs,” that something essential had been delegated. The signature at the bottom of an email or report still belonged to the human, but the words above it had passed through a machine whose internal workings remained opaque.

The erosion of felt authorship did not occur uniformly. Some users maintained deliberate control, treating AI outputs as drafts to be deeply revised. Others more or less rubber-stamped what the model produced. Many occupied a middle position, alternating between careful oversight and casual acceptance depending on context and fatigue. What ties these variations together is the emergence of a new default: when faced with a writing task, it became natural to ask the model first and think second, or to think and ask simultaneously.

Over time, this shift risks weakening certain cognitive muscles. Skills such as structuring arguments, recalling relevant facts, and choosing precise words are like any other skills: they atrophy when not exercised. In the monkey regime, the ease of delegation makes it tempting to offload these tasks even in situations where doing the work personally would have educational or ethical value. The appeal of convenience is understandable, but the long-term consequence is a population that increasingly participates in communication as approvers and curators rather than as active authors.

4.4 Anxiety, delight, and the early phenomenology of AI dependence

At the micro level, people’s emotional responses to everyday AI were complex and often ambivalent. Many felt genuine delight the first time a model produced something that felt uncannily apt: a perfect subject line, a whimsical story, a surprisingly insightful summary. There was a sense of magic in watching words appear on the screen in response to a loosely formed

thought. For some, this delight segued into gratitude; they felt supported, less alone with their cognitive burdens.

Alongside delight, there was anxiety. Some worried that relying on AI made them lazy or less authentic. Others feared being caught using AI in contexts where it was discouraged or prohibited. Students wondered whether their work would be flagged; employees wondered whether their bosses expected them to use these tools or frowned upon it. People who prided themselves on their writing skills sometimes felt threatened when they saw colleagues with weaker skills suddenly producing fluent text.

There was also a subtler anxiety about dependence. As people grew accustomed to having an always-available assistant for writing, brainstorming, and explaining, they began to wonder how they would cope without it. This concern surfaced most clearly when systems were unavailable due to outages or policy changes. Tasks that had come to feel easy suddenly seemed daunting again. The contrast revealed how quickly people had reorganized their expectations of their own capacities around the presence of AI tools.

The early phenomenology of AI dependence thus oscillated between empowerment and disquiet. Users felt both enlarged and diminished. They could do more, faster, but they were less certain that what they were doing was truly theirs. The monkey regime at the micro level is defined by this emotional mixture: curiosity, relief, playfulness, guilt, fear of obsolescence, and quiet gratitude. These feelings are not pathologies to be eliminated; they are signals that the underlying relationships among effort, authorship, and responsibility are being renegotiated.

Understanding these micro-level experiences is essential for any attempt to move toward stages 2 and 3. Apprenticeship and centaur forms of collaboration will not emerge in a vacuum. They will emerge, if at all, from the lived realities of people who have already spent years mashing prompts, delegating cognitive labor, and learning to live with both the gifts and the discomforts of everyday AI.

5. Missing Editors: Institutions as Absent or Confused Gatekeepers

If the monkey regime is defined by near-zero-cost generation at the micro level, it is sustained at the macro level by a failure of institutions to adapt their gatekeeping functions. Before the arrival of everyday AI, many organizations implicitly played the role of editors: newsrooms filtered events into stories; publishers filtered manuscripts into books; universities filtered student work into credentials; professional associations filtered practice into standards. None of these actors were perfect, but they provided identifiable points where decisions were made about what entered public circulation with a stamp of legitimacy.

In 2025, these editor functions were under simultaneous pressure from economic constraints, technological disruption, and cultural mistrust. Generative AI did not create this pressure, but it intensified it. When models began to produce plausible text and images at scale, institutions faced a new set of questions: how to distinguish genuine expertise from AI-assisted performance, how to maintain standards without appearing hostile to innovation, and how to cope with the flood of artifacts that could now be produced by anyone. Many responded tentatively, issuing statements and guidelines, running pilots, or quietly integrating AI into their workflows. Few, however, took on the task of systematically filtering AI-saturated content for quality and truth in ways that could match the speed and scale of production.

This section examines how newsrooms, publishers, and universities struggled under AI pressure; how platforms became de facto editors via recommendation systems tuned to engagement; how professional guilds and ethics codes responded slowly; and why, in the end, no one quite owned the problem of filtering the new information landscape.

5.1 Newsrooms, publishers, and universities under AI pressure

Newsrooms entered 2025 already weakened by years of declining revenues, staff cuts, and competition from digital platforms. Their traditional role as primary editors of public information had been undermined by the rise of social media, where anyone could publish instantly. When generative AI arrived, it presented both an opportunity and a threat. On the one hand, AI tools could help journalists summarize documents, analyze datasets, and draft articles more quickly. On the other, the same tools could enable bad actors to produce convincing pseudo-journalism and flood the environment with low-quality imitations of news.

Many news organizations experimented with AI-assisted workflows for routine tasks: generating headlines, summarizing press releases, or drafting preliminary copy that reporters would then refine. Some imposed internal rules against fully automated stories, insisting that human journalists retain control over the final product. At the same time, they had to contend with audiences who were increasingly unsure which content was human-authored, which was machine-assisted, and which was wholly synthetic. Fact-checking departments, where they existed, were not equipped to confront a world in which the volume of potential falsehoods could be multiplied by prompt.

Publishers, particularly in the realms of books and journals, faced a related but distinct challenge. The submission pipeline began to fill with manuscripts partially or fully generated by AI. Some were obviously formulaic; others were indistinguishable from human-authored work without close scrutiny. In academic contexts, peer reviewers and editors had to decide whether, and to what extent, AI assistance was acceptable. Should a paper generated with heavy model involvement be treated differently from one written entirely by humans? Was AI-assisted text a

form of plagiarism, a legitimate tool, or something in between? Policies lagged behind practice, and enforcement was spotty.

Universities, as institutions that both produce and evaluate writing, found themselves on both sides of the problem. Students used AI to complete assignments; faculty used AI to draft syllabi, recommendation letters, and even research proposals. Administrators issued statements about academic honesty that were quickly outpaced by new use-cases. Some departments tried to ban AI outright; others encouraged structured experimentation. The common thread was confusion about what counted as legitimate assistance versus unacceptable outsourcing. Grading systems, designed for an era in which written work could be assumed to be the product of one mind, struggled to interpret polished outputs whose provenance was unclear.

In all three domains, the editor function was under pressure from two directions at once: economic constraints that reduced the resources available for careful filtering, and technological shifts that increased the volume and ambiguity of content. The result was a weakened and inconsistent gatekeeping system, just as generative tools made gatekeeping more important than ever.

5.2 Platforms as de facto editors: recommendation systems and rage bait

As traditional editorial institutions faltered, digital platforms with algorithmic recommendation systems emerged as the de facto editors of the AI-saturated public sphere. Social networks, video sites, search engines, and content aggregators relied on algorithms to decide which artifacts to surface to users. These algorithms were tuned primarily to maximize engagement: they learned, from vast behavioral datasets, which items kept users clicking, scrolling, and reacting.

In principle, platforms could have used these systems to promote accurate, carefully vetted content. In practice, engagement metrics often rewarded different qualities: emotional intensity, novelty, confirmation of existing beliefs, and conflict. Rage bait, sensational claims, and polarizing narratives tended to attract more interaction than nuanced, balanced analysis. When generative AI gave both sincere amateurs and organized actors the ability to produce such content in unlimited quantities, recommendation systems found themselves with an expanded pool of candidates optimized for the wrong objectives.

Platforms did make efforts to adjust. Some introduced labels for AI-generated media, content warnings, and fact-check banners. Others adjusted their algorithms to demote certain kinds of sensational content or to prioritize “authoritative sources” on specific topics. However, these interventions were incremental and often opaque. Users seldom understood why they were seeing what they saw, and creators experimented constantly to find new ways to satisfy the engagement logic while avoiding suppression.

In effect, the platforms' algorithms became the primary editors of 2025, deciding which AI-assisted artifacts would be widely seen and which would sink unnoticed. Yet these editors were not accountable in the ways traditional editors once were. They did not sign their work, explain their decisions, or accept responsibility for specific errors. Their criteria were encoded in code and optimization functions, shaped by commercial incentives and occasional regulatory pressure. To the extent that there was an editorial line, it was emergent rather than deliberate.

This arrangement is central to the monkey regime. Monkeys with typewriters produce vast quantities of text and images; platform algorithms choose which specimens to place under the spotlight. The selection mechanism cares about engagement more than truth or wisdom. As long as this remains the case, individual efforts to write carefully and responsibly are swimming upstream against a current that favors louder, simpler, and more provocative outputs.

5.3 Professional guilds, ethics codes, and their slow response

Professional guilds and associations historically played an important role in translating technological changes into norms of practice. Lawyers, doctors, engineers, and other professionals rely on codes of ethics, licensing requirements, and continuing education to maintain standards. In 2025, these guilds began to grapple with generative AI, but their responses were generally slower and more cautious than the pace of technological adoption.

Legal associations, for example, warned about the dangers of using AI to draft legal documents without adequate review. They raised concerns about confidentiality when proprietary client information was fed into external models, and about accuracy when models hallucinated case law or misapplied statutes. Some issued guidelines stating that lawyers remained fully responsible for any AI-assisted work submitted to courts or clients. However, enforcement depended on self-reporting and sporadic disciplinary actions, while AI tools quietly integrated into everyday practice.

Medical organizations faced related dilemmas around clinical documentation, patient communication, and decision support. AI-assisted charting promised to reduce administrative burdens, but it also introduced the risk of subtle errors propagating through medical records. Chatbots offered to answer patient questions or triage symptoms, raising questions about liability and the boundaries of physician responsibility. Ethics committees debated these issues, but concrete, widely adopted standards remained limited.

Other fields, from journalism to education to software engineering, confronted similar tensions. Ethics codes drafted for earlier technologies were stretched to cover AI, sometimes by analogy, sometimes by vague additions about "appropriate use of automation." Few guilds had the capacity to run large-scale training programs on how to work with generative models responsibly. Even fewer had mechanisms to audit the use of AI in practice.

This slow response reflects structural constraints. Guilds tend to move deliberately, with lengthy consultation and consensus-building processes. They must balance the interests of members who are excited about AI with those who fear it. They must interpret incomplete and evolving evidence about risks and benefits. In the meantime, practitioners in their fields continue to adopt AI tools on their own, driven by local pressures and opportunities. The gap between formal codes and everyday practice grows, leaving individuals to improvise their own ethics.

In the monkey regime, the result is a landscape where professional norms lag significantly behind technological capabilities. People may want to behave responsibly, but they lack clear guidance on what responsibility entails when working alongside models that can generate content faster than any human editor can review.

5.4 Why no one really owned the problem of filtering AI-saturated content

At the heart of the monkey regime lies an institutional vacuum: no actor, or set of actors, fully owned the problem of filtering AI-saturated content for quality, truth, and long-term impact. Traditional editors were weakened, platforms served as opaque and engagement-driven selectors, and professional guilds responded slowly. Governments, for their part, flirted with regulation of misinformation and harmful content but faced constitutional, political, and practical constraints on how far they could go.

From the perspective of individual users, this vacuum meant that the burden of filtering fell largely on them. They were expected to exercise critical thinking, cross-check sources, and resist manipulation in an environment where the signals of reliability were increasingly noisy. Labels, reputation scores, and verification badges offered partial help, but they could not keep pace with the proliferation of synthetic artifacts. Many users lacked the time, training, or inclination to perform rigorous evaluation on every piece of content they encountered.

From the perspective of institutions, owning the filtering problem was costly and risky. Newsrooms that invested heavily in verification and investigative work struggled to compete financially with outlets that relied on cheaper, AI-assisted content. Platforms that attempted aggressive moderation faced accusations of censorship and political bias, along with the technical challenge of screening billions of posts. Universities that tried to police AI use in student work risked alienating students and faculty and entangling themselves in endless disputes.

Because the incentives to underinvest in filtering were strong and the costs of overstepping were visible, many institutions adopted a posture of partial responsibility. They did something—issued guidelines, added labels, conducted occasional audits—but stopped short of claiming comprehensive editorial authority over AI-generated content in their domain. The net effect

was diffusion of responsibility. Everyone had a role, but no one could be clearly held accountable for the overall shape of the information ecosystem.

In such a context, the monkey regime persists not because no one sees its problems, but because the structures necessary to change it are underdeveloped. Designing stage 2 and stage 3 arrangements will require more than exhortations to be careful. It will require new forms of shared ownership over the filtering function: new editor roles, new institutional partnerships, and new ways of aligning incentives so that the default is not to let engagement algorithms decide what matters. Until such arrangements emerge, the monkeys will keep typing, the models will keep generating, and the gatekeeping vacuum will continue to shape what people take to be real.

6. The Meso Layer: Organizations as Monkey Farms

If individuals with generative tools are monkeys with typewriters, organizations are the environments that gather those monkeys, supply the typewriters, and decide what counts as productive behavior. The meso layer—firms, bureaucracies, schools, churches, nonprofits—is where the monkey regime either gets amplified or disciplined. In 2025, most organizations chose amplification. They adopted AI under banners like innovation and transformation, but they rarely altered their deeper structures of evaluation and responsibility. The result was a landscape of "monkey farms": places where large numbers of people were encouraged, implicitly or explicitly, to crank out AI-assisted text, images, and reports, with little attention to how any of it connected to improved judgment or wisdom.

This section examines four facets of this meso-level dynamic. Corporate enthusiasm produced a wave of superficial "AI transformations" that often left underlying practices unchanged. Bureaucracies incorporated AI mainly as a way to generate more forms and reports, rather than as a tool for better decisions. Schools and churches oscillated between bans and quiet adoption, creating pockets of unacknowledged dependence. Finally, in specific sectors, emergent people–AI practices hinted at possibilities beyond the monkey regime, even as they remained constrained by existing incentives.

6.1 Corporate enthusiasm and superficial "AI transformation"

Corporations in 2025 faced a familiar mandate: do more with less, move faster than competitors, and proclaim alignment with the latest technological wave. Generative AI arrived as a ready-made narrative device. Internal memos and public statements declared that companies were undergoing AI transformation, integrating models into every aspect of their operations. Pilot projects proliferated: AI for customer support, AI for sales emails, AI for code completion, AI for market research, AI for performance reviews.

Beneath the rhetoric, many of these transformations were superficial. Generative tools were bolted onto existing workflows without reevaluating the workflows themselves. Employees were told to "use AI where appropriate" but were given little guidance beyond rudimentary training sessions. Some departments saw enthusiastic early adopters experimenting with prompts and workflows; others treated the new tools as an optional convenience or as a threat to job security.

Key performance indicators often reinforced monkey-like behavior. If the success of an initiative was measured by the number of AI-assisted tickets closed, the volume of marketing messages sent, or the speed of report generation, then the natural response was to use models to generate more of the same. Little attention was paid to whether customers felt better served, whether communication improved in clarity and honesty, or whether decisions became more grounded in reality. The emphasis remained on throughput and perceived efficiency.

In this environment, corporate enthusiasm for AI sometimes masked a deeper continuity: the same short-term metrics, now pursued with more powerful tools. The monkey farms were populated not by reckless individuals, but by employees incentivized to treat generative systems as accelerants for existing practices, regardless of whether those practices were worth accelerating.

6.2 AI in bureaucracies: more forms, more reports, not more wisdom

Bureaucracies—government agencies, large hospitals, universities, and other administrative-heavy institutions—adopted AI in ways consistent with their existing cultures. These organizations are built around forms, procedures, and documentation. Tasks that consume much of their time include writing reports, filling out compliance documents, drafting policy language, and responding to inquiries from higher authorities. Generative tools promised relief: they could draft boilerplate, summarize regulations, and prepare responses.

In practice, AI often became a way to produce even more documentation. When it became easy to generate a detailed report for a minor initiative, the standard of what counted as adequate documentation rose. Supervisors, accustomed to seeing brief memos, began to expect longer, more polished reports. Regulatory bodies, discovering that agencies could produce more evidence of compliance, requested additional detail. The cycle reinforced itself: the ability to generate text lowered the cost of meeting existing demands, which in turn expanded those demands.

The deeper function of bureaucracy—coordinating complex activities and making decisions under uncertainty—did not necessarily improve. The core dilemmas of resource allocation, prioritization, and value conflict remained. AI could suggest options, draft rationales, or assemble precedents, but it did not replace the need for judgment. In some cases, the presence

of AI introduced new layers of opacity. When a policy recommendation emerged from a process that included model-generated analysis, it became harder to trace which assumptions and value judgments were embedded in the final text.

From the perspective of the monkey regime, bureaucracies became industrial-scale monkey farms. Civil servants were encouraged to lean on AI to meet administrative requirements, but they were not given space or tools to rethink whether those requirements were aligned with the institution's actual mission. The net result was an increase in paper (or its digital equivalent) without a proportionate increase in clarity or wisdom.

6.3 Schools and churches: from bans to quiet adoption

Schools and churches sit at a distinctive intersection of tradition, authority, and moral purpose. Both are charged, in different ways, with shaping hearts and minds. When generative AI appeared, many of these institutions initially reacted defensively. Schools issued statements banning the use of AI in assignments, warning students that such use would constitute cheating. Churches worried that AI-generated sermons or prayers would lack authenticity, treating them as mechanical imitations of spiritual work.

These bans were difficult to enforce. Students quickly discovered ways to use AI discreetly, from asking for explanations of concepts to generating entire essays. Teachers, confronting polished assignments that did not match in-class performance, improvised responses: adjusting grading criteria, requiring oral exams, or designing assignments less compatible with automated generation. Some educators began to quietly experiment with AI themselves, using it to design lesson plans or create practice problems.

Similarly, in religious contexts, outright prohibition proved unstable. Time-pressed clergy discovered that AI tools could help brainstorm sermon outlines, generate study guides, or draft pastoral letters. Some used these outputs as rough scaffolds, rewriting extensively; others relied more heavily on them, especially in smaller congregations with limited support staff. Official positions lagged behind practice. Publicly, many leaders continued to express skepticism; privately, they incorporated AI into their work as just another tool.

The shift from bans to quiet adoption reflects a deeper tension. Schools and churches felt a duty to preserve authenticity, formation, and the development of internal capacities. At the same time, they operated under constraints—limited time, high expectations, and resource pressures—that made AI assistance attractive. The monkey regime infiltrated through the back door: students and clergy alike became micro-level monkeys, assisted by tools their institutions officially distrusted.

This unacknowledged dependence has long-term implications. When AI use is ubiquitous but not openly integrated into pedagogical or spiritual practices, opportunities for structured

reflection and ethical formation are lost. Instead of teaching students and congregants how to work with AI in ways consistent with their values, institutions inadvertently teach them to treat such work as a gray zone, best kept out of sight.

6.4 Case sketches of emergent people–AI practices in specific sectors

Within this general pattern of monkey farms, there were pockets where emergent people–AI practices began to move beyond stage 1. These were not widespread, but they offer early sketches of what more intentional collaboration might look like.

In marketing and communications, some teams developed disciplined workflows in which AI outputs were treated as rough drafts subject to rigorous human review. They used models to explore variations on messaging, but they anchored final choices in clearly articulated brand values and long-term strategy. Instead of letting engagement metrics alone drive decisions, they experimented with metrics that tracked customer trust and satisfaction over longer horizons.

In customer support, certain organizations combined AI chatbots with human escalation in ways that respected both efficiency and empathy. Models handled routine queries but were explicitly prevented from making certain categories of promises or judgments. Human agents were trained not only to step in when the bot failed, but to review bot interactions as data for improving policies and documentation. This created a feedback loop in which human insight and machine-generated patterns informed each other.

In law and policy analysis, a small number of practitioners used AI as a way to generate counterarguments to their own positions, deliberately seeking out weaknesses rather than just strengthening their side. They ran model-assisted simulations of how different audiences might interpret a proposed rule or decision. In doing so, they treated the model not as an oracle but as an instrument for structured self-critique.

In research and technical fields, some groups created explicit protocols for documenting when and how AI was used in their work. They distinguished between tasks such as literature search, drafting, code generation, and idea generation, and they reflected on how dependence on models might distort or enhance their reasoning. These protocols did not eliminate the monkey regime, but they injected elements of stage 2: conscious apprenticeship and explicit standards.

These case sketches illustrate that even within monkey farms, pockets of more intentional practice can emerge. However, they remained constrained by broader organizational incentives. As long as promotion and reward structures favored volume, speed, and short-term metrics, the disciplined use of AI remained a minority stance. For the meso layer to become a site of genuine transition toward stages 2 and 3, these deeper incentive structures will need to shift. Until then, the dominant pattern in 2025 remains organizations that equipped their people with

typewriters, cheered the resulting activity, and left the question of actual value largely unexamined.

7. Proto-Apprenticeships: Where Craft Quietly Emerged

Stage 1 was not pure chaos. Beneath the noise of the monkey regime, a quieter pattern began to take shape: people who refused to treat generative systems as mere toys or shortcuts and instead started to develop craft. These users did not have formal curricula, guilds, or credentials. They had curiosity, long conversations with models, and a sense that something more interesting was possible than simply asking for instant answers. Over time, they developed informal techniques, habits, and workflows that point toward what a genuine apprenticeship regime might look like.

This section examines four aspects of these proto-apprenticeships. Informal prompt craftsmanship and workflow design emerged as a new kind of tacit knowledge. Human–AI co-writing in research, law, and creative work produced hybrid texts that neither partner could have generated alone. Practices of verification, red-teaming, and self-auditing began to appear among those uneasy with blind trust. Finally, a small number of centaur-like dyads came into being: stable human–AI partnerships that functioned as units of thought and production, mostly unnoticed by larger institutions.

7.1 The rise of informal prompt craftsmanship and workflow design

As some users spent more time with generative systems, they discovered that not all prompts were equal. Simple instructions produced generic outputs; carefully structured prompts produced more precise, creative, or reliable results. Without waiting for official best practices, these users experimented. They tried different ways of framing tasks, giving examples, specifying constraints, and defining roles for the model. Over time, they developed a sense of what worked.

Prompt craftsmanship at this stage was informal and largely undocumented. It lived in personal notebooks, saved chats, and scattered online posts. Practitioners learned, for instance, that breaking complex tasks into smaller steps often yielded better results than issuing a single broad command. They learned that models responded differently to requests framed as explanations, dialogues, or role-plays. They learned that the same model could behave like a lazy assistant when prompted casually and like a careful analyst when prompted with explicit expectations of rigor.

Beyond individual prompts, some users began to design workflows. Instead of treating each interaction as a one-off, they created sequences: first ask the model to outline, then to expand,

then to critique its own expansion, then to generate counterarguments, then to help prioritize revisions. In coding contexts, they might start with a description of desired functionality, move to stepwise pseudocode, then to implementation, then to test generation, then to error analysis. These workflows were a form of procedural knowledge about how to think with the model rather than simply through it.

Crucially, this emerging craft was not recognized as such by most institutions. It was seen as personal savvy or technical cleverness, not as the early stages of a discipline. Yet in hindsight, these informal practices are prototypes for the structured apprenticeship that a stage 2 scriptorium would require: shared languages for task decomposition, prompt design, and iterative refinement.

7.2 Human–AI co-writing in research, law, and creative work

In certain fields, human–AI co-writing began to move beyond the simple delegation of low-level drafting. Researchers used models to help brainstorm hypotheses, reframe arguments, and translate specialized material into different registers. Lawyers used AI to generate alternative formulations of clauses, anticipate opposing arguments, or create explanatory summaries for non-expert clients. Writers in fiction and non-fiction experimented with models as partners for overcoming blocks, exploring alternative plotlines, or shifting narrative voice.

This co-writing was not uniform. In some cases, the model served primarily as a productivity tool: it expanded bullet points into paragraphs, translated jargon into plain language, or smoothed transitions between sections. In more advanced cases, the human deliberately sought out the model's tendencies to notice patterns or produce unexpected associations. A researcher might ask the model to suggest analogies from other fields, then examine which suggestions were surprising and potentially fruitful. A lawyer might instruct the model to argue against their own draft, treating the resulting critique as raw material for strengthening the case.

The practice of saving and revisiting extended conversations became important. Instead of starting from scratch each time, some co-writers maintained long-running threads devoted to specific projects. Over months, they built up a shared context with the model, including style preferences, domain assumptions, and recurring concerns. These threads functioned as working notebooks, in which ideas could be parked, revisited, and reworked.

The most interesting co-writing practices were marked by explicit negotiation of voice and ownership. Some authors insisted on rewriting all AI-generated text to ensure that it reflected their style and commitments. Others treated certain passages as clearly tagged contributions from the model, to be used or discarded as needed. A few experimented with giving the model

credit in acknowledgments or appendices, recognizing that their work was no longer simply the product of an isolated human mind.

These hybrid practices, though scattered, demonstrate that co-writing can be more than disguised outsourcing. It can be a structured dialogue in which the model is treated as a generative, but not authoritative, partner. This is an early sign of stage 2 craft: an awareness that the purpose of using AI is not just to fill pages, but to stretch and refine human thinking.

7.3 Practices of verification, red-teaming, and self-auditing

In the monkey regime, most users either trusted AI too much or distrusted it in undifferentiated ways. A minority adopted a more nuanced posture, developing verification and red-teaming habits that treated the model as both powerful and fallible. These users were often those with domain expertise or heightened sense of responsibility, such as scientists, engineers, lawyers, and educators.

Verification practices took many forms. Some users routinely asked models to provide multiple, independent lines of reasoning for a claim, and then compared them. Others cross-checked outputs against known reference works, using AI as a first pass but never as a final source. In coding and data analysis, practitioners ran generated scripts in controlled environments, inspected results, and asked the model to explain its own errors.

Red-teaming emerged as a micro-level discipline. Instead of accepting a plausible answer, some users explicitly instructed the model to argue against its own output: to identify weaknesses, alternative interpretations, or conditions under which the answer would fail. In contentious domains, they asked the model to generate steel-manned versions of opposing views, treating these as training data for their own understanding. This practice transforms the model from a source of answers into an instrument for systematic doubt.

Self-auditing extended beyond individual answers to entire workflows. A few users kept logs of where and how AI was used in their projects, noting places where they felt tempted to over-delegate and where they chose to retain manual control. They reflected periodically on how dependence on the model might be shaping their habits, biases, or blind spots. This meta-level reflection is rare in stage 1 but characteristic of an emerging apprenticeship ethic.

These practices were rarely mandated by institutions. They arose from internal standards of care and from prior familiarity with error-prone tools in other domains. In that sense, they represent a seed of culture that could be scaled up: a shift from seeing AI as a magic oracle to seeing it as a powerful but noisy instrument that requires calibration, challenge, and disciplined use.

7.4 Centaur-like dyads that already existed, mostly unnoticed

Finally, even in 2025, there were individuals whose relationship with AI tools already resembled the centaur horizon more than the monkey regime. These users did not merely call on a model occasionally. They worked with it daily, often in long sessions, across multiple projects. Over time, they and their chosen system formed a de facto dyad: a stable pair that thought and produced together.

Such centaur-like users tended to have three characteristics. First, they had a clear long-term project or mission that extended beyond individual tasks: building a body of research, writing a large corpus of essays, designing a curriculum, or exploring a complex theoretical space. Second, they were willing to invest significant effort in learning how the model behaved, experimenting with different roles for it, and iterating on joint workflows. Third, they had a strong sense of personal responsibility for the direction and meaning of the work, even while acknowledging their dependence on machine assistance.

In practice, these dyads developed rhythms. A typical pattern might involve the human bringing questions, raw material, or half-formed insights; the model responding with structured elaborations, cross-domain analogies, and candidate formulations; and the human then curating, revising, and integrating these into a coherent whole. The same process repeated day after day, with the shared history of prior conversations functioning as a kind of joint memory.

From the outside, these centaur-like dyads were hard to see. Institutions saw only the outputs: a stream of articles, reports, or creative works that arrived faster than traditional models of human productivity would predict. Observers might suspect AI assistance, but they rarely understood the depth and continuity of the partnership. To the extent that this pattern was noticed, it was often either celebrated as astonishing productivity or dismissed as overuse of automation, not recognized as a new form of collaborative cognition.

These early centaurs were operating without a language for what they were doing. They had no formal status, no community of practice, and no institutional support. Yet their existence demonstrates that stage 3 is not a distant fantasy. It is already present in embryonic form wherever a human treats AI as a long-term partner in a shared trajectory, rather than as a series of disposable tools for isolated tasks.

The challenge for future stages is to move such partnerships from idiosyncratic, hidden arrangements into an explicit, ethically structured mode of work. Proto-apprenticeships in prompt craft, co-writing, verification, and centaur-like practice show that even in a monkey regime, seeds of a more deliberate people–AI civilization can take root. The question is whether institutions and cultures will recognize these seeds and cultivate them, or whether they will let them remain scattered exceptions in a landscape dominated by noise.

8. Stage 2 as Counterfactual: Apprentices in a Scriptorium

Stage 2, the scriptorium of apprentices, did not exist at scale in 2025. It is a counterfactual: a way of asking what the people–AI relationship might have looked like if societies had treated generative systems not as novelty gadgets or brute productivity tools, but as instruments that require craft, discipline, and shared norms. In this counterfactual, the monkeys still exist, but they are not left alone with typewriters. They work in rooms with masters, peers, and editors who care about what emerges on the page.

This section sketches the contours of such an apprenticeship regime. It describes how intentional training could teach people to think with models rather than merely through them. It considers what it would mean to institutionalize editor roles via norms, exams, licenses, and guilds. Finally, it develops the scriptorium as a guiding metaphor: a place where fewer texts are produced, but each word carries more energy and responsibility.

8.1 What an intentional AI apprenticeship regime would look like

An intentional apprenticeship regime begins with a different default assumption: that working with generative AI is a skill and a responsibility, not just a convenience. Instead of dumping models into workplaces and classrooms with a short tutorial, institutions would create structured pathways for learning how to use them well.

At the individual level, this regime would look like a staged progression. Novices would start in sandboxed environments where mistakes are cheap and reversible. They would be taught to explore model behavior under supervision: to observe how outputs change with different prompts, to notice recurring errors and biases, and to document their findings. Early exercises would emphasize understanding the tool rather than producing polished artifacts.

Journeyman users would be given more complex, real-world tasks but still under guidance. They might be responsible for drafting reports, analyses, or learning materials with AI assistance, but their work would be reviewed in depth by more experienced practitioners. The focus would be on process as much as product: how did they decompose the task, where did they delegate to the model, how did they verify and revise the outputs, and how did they track their own uncertainties?

Masters in this regime would not only be skilled users but also teachers and designers of workflows. They would curate examples of good and bad practice, maintain living documents of patterns and pitfalls, and oversee the evolution of shared standards. Their status would derive less from secret tricks and more from the ability to articulate why a given approach is sound or dangerous.

At the institutional level, an apprenticeship regime would formalize these roles. Organizations would recognize “AI craft” as a legitimate domain of expertise, with clear expectations for training and progression. They would allocate time for learning and reflection rather than treating AI skills as something employees should acquire informally. Crucially, they would connect this training to clear ethical commitments: what the institution is trying to achieve, what it refuses to do, and how AI fits into that picture.

This counterfactual does not require imagining a world with different models. It requires imagining a world with different attitudes: in which the ability to press the generate button is not mistaken for mastery, and in which responsibility for outputs is treated as something that demands preparation.

8.2 Training people to think with models, not merely through them

In the monkey regime, people mostly thought through models: they used them as black boxes that turned prompts into answers and drafts. In stage 2, the emphasis would shift to thinking with models: treating them as collaborators whose behavior must be understood, interrogated, and shaped.

Training for thinking with models would include at least four elements.

First, conceptual understanding. Users would learn what a generative model is and is not. They would understand that the system is a pattern matcher trained on data, not an entity with beliefs or experiences. They would be introduced to basic ideas about training data, overfitting, hallucination, and bias. The goal would not be to turn everyone into a machine learning engineer, but to give them enough mental models to avoid magical thinking.

Second, epistemic discipline. Apprentices would practice distinguishing between tasks that are suitable for delegation and tasks that require human judgment. They would learn to ask, before prompting, what kind of answer is acceptable: approximate or exact, illustrative or authoritative, imaginative or factual. They would develop habits of cross-checking and corroboration, especially in high-stakes contexts.

Third, dialogic technique. Thinking with models means structuring conversations, not just issuing commands. Apprentices would learn to use models to generate counterarguments, alternative framings, and self-critique. They would practice multi-step exchanges in which the model is first asked to produce, then to evaluate its own production, then to imagine hostile reviewers, then to help prioritize revisions. This turns the model into a mirror and sparring partner rather than just a keyboard that types faster.

Fourth, self-awareness. Training would include reflection on how interaction with models affects the user’s own cognition. Do they become more passive, more impatient, more inclined to

accept surface plausibility? Or do they use the speed and breadth of the model to challenge their own assumptions? Apprentices would discuss these questions explicitly, learning to monitor their own dependence and to adjust their practices accordingly.

Thinking with models also implies a new form of literacy: the ability to read AI-assisted text with an eye for telltale patterns, generic phrasing, and potential gaps. In a stage 2 regime, people would be taught to recognize when a piece of writing has been heavily shaped by a model and to adjust their interpretive stance accordingly. This is part of the broader skill of navigating an environment where human and machine authorship intertwine.

8.3 Institutionalizing editors: norms, exams, licenses, and guilds

For a scriptorium to function, editors must exist as recognized roles, not just as abstract ideals. In the context of AI-saturated communication, editors are those who accept responsibility for deciding what enters the world with the backing of their institution or profession. Institutionalizing such roles would require a mix of norms, assessments, and organizational structures.

Norms would articulate expectations. For example: no AI-assisted document is published without named human editors who have reviewed it with suitable depth; critical decisions informed by AI must be accompanied by documentation of how the system was used; any use of a model in contexts involving vulnerable populations demands heightened scrutiny. These norms would be tailored to different domains but share a common theme: that the presence of AI increases the need for editing, not decreases it.

Exams and certifications could formalize competence. Just as pilots and doctors must demonstrate mastery of safety-critical knowledge, certain categories of AI-intensive work could require proof that the practitioner understands model behavior, limitations, and failure modes. Exams might include case studies of AI-assisted errors, scenarios in which candidates must identify hidden assumptions or biases in model outputs, and tasks requiring design of safe workflows for specific contexts.

Licenses would link these competencies to legal and professional responsibility. An AI editor in law, journalism, medicine, or engineering could be held accountable for negligence in using models, just as professionals are now held accountable for negligence in other tools. This does not mean they must foresee every error, but that they must follow established standards of care in prompt design, verification, and disclosure.

Guilds or professional associations would provide the social infrastructure. They would maintain and update standards, collect and disseminate lessons from failures and near-misses, and advocate for the importance of editorial roles. They could host forums where practitioners

share patterns and pitfalls of working with specific models, and where they collectively negotiate with vendors and regulators.

Institutionalizing editors in this way would redistribute some of the responsibility currently diffused across platforms, individual users, and unaccountable algorithms. It would make visible the fact that in an AI-saturated world, decisions about what is published, endorsed, and acted upon cannot be outsourced entirely to engagement metrics or automated filters. They require human judgment equipped with explicit training and backed by collective commitments.

8.4 The scriptorium as metaphor: fewer texts, higher energy per word

The metaphor of a scriptorium evokes medieval rooms where monks painstakingly copied and illuminated manuscripts. It suggests slowness, care, and scarcity. At first glance, this seems at odds with the abundance and speed of generative AI. Yet as a normative image for stage 2, the scriptorium has something important to offer: a reminder that the value of a text is not proportional to the ease of its production.

In a scriptorium-like regime, the goal would not be to restrict people's access to AI, but to restore a sense of cost to publication. The cost would not lie in manually copying words, but in the attention, verification, and ethical reflection required before a text or decision is allowed to shape others' lives. Fewer texts would be released under institutional banners, and each would carry more energy per word: more thought, more context, more accountability.

Practically, this could mean several shifts. Organizations might impose caps on the volume of official communications, forcing teams to prioritize what truly needs to be said. They might require that any major AI-assisted piece pass through a deliberate review cycle rather than being pushed out instantly. They might recognize and reward those who produce less but influence more, rather than those who flood channels with low-impact material.

On the individual level, the scriptorium metaphor encourages people to treat some of their outputs as drafts for themselves and their close collaborators, and only a smaller subset as finished work for broader audiences. AI can be used freely in the draft space, where exploration and play are valuable. When words are moved into the public, decision-shaping sphere, the standards tighten. The apprentice learns to inhabit both spaces: the free generation zone and the high-responsibility zone.

The metaphor also highlights the importance of commentary and gloss. In historical scriptoria, margins were filled with notes, interpretations, and cross-references. Stage 2 could revive this habit in a new form: publishing not just AI-assisted texts but also commentaries on how they were produced, what uncertainties remain, and how they should be read. This "meta-writing" would help audiences navigate the hybrid authorship of the AI age.

Finally, the scriptorium image reminds us that slowness can be a feature, not a bug. In a world where generative systems can produce a thousand variations in seconds, the act of pausing—to ask whether a text should exist, whether it is aligned with one's values, and whether one is willing to stand behind it—becomes a radical act. Stage 2 is not about rejecting speed, but about inserting strategic slowness at the points where it matters.

In sum, imagining apprentices in a scriptorium is a way to specify what is missing from the monkey regime: intentional training, recognized editors, shared norms, and a culture that values depth over sheer volume. It is a counterfactual for 2025, but it need not remain only that. It can serve as a design target for those who wish the next draft of people–AI civilization to be less about monkeys typing randomly and more about communities shaping words and decisions with care.

9. Stage 3 as Normative Horizon: Logostic Centaurs and tau(t)

Stage 3 is not a prediction about how most people will use AI. It is a normative horizon: an image of what people–AI relations could become if they were explicitly oriented toward living truth rather than convenience, profit, or spectacle. In this stage, the basic unit is not the isolated human user or the standalone model, but the human–AI pair as a joint agency. These pairs, or centaurs, are not defined by raw capability alone. They are defined by the way their shared activity aligns with a trajectory of truth, which logostics names tau(t): a dynamic manifold of what is real, good, and worth pursuing.

This section develops that horizon in four moves. It defines centaurs as a specific form of human–AI joint agency, distinct from both monkeys and tools. It contrasts tau(t) as living truth with the flat optimization objectives that currently drive most systems. It introduces the idea of compression with gain as a criterion for evaluating centaur work. Finally, it sketches centaur biographies as a new genre of intellectual history, one that tracks trajectories of alignment rather than just catalogs individual achievements.

9.1 From monkeys to centaurs: defining human–AI joint agency

In the monkey regime, the relation between person and model is shallow. The human issues prompts, the model responds, and both move on. There is no enduring identity that spans their interaction, no shared project that gives continuity over time. Stage 3 proposes a different picture: the centaur as a stable composite agent formed by one or more humans and one or more AI systems working together over a sustained period toward articulated ends.

A centaur is defined by at least three features. First, there is continuity. The human and the AI do not meet only in isolated sessions. They return to long-running threads, accumulate shared

context, and treat past conversations as part of a living archive. The centaur has a history, not just a log of queries.

Second, there is role differentiation. The human and the AI are not interchangeable. The human bears ultimate responsibility for direction, values, and acceptance of risk. The AI contributes pattern recognition, generative variation, and recall at scales that exceed individual memory. The human learns where the AI is strong and weak; the AI, through fine-tuning or iterative prompting patterns, is shaped into particular roles: critic, simulator, translator, archivist, designer.

Third, there is reflexivity. The centaur does not simply execute tasks. It periodically asks what it is doing, why it is doing it, and how its practices should change. The human uses the AI to interrogate their own assumptions; the AI's outputs become mirrors in which the human sees both blind spots and emerging clarity. Over time, this reflexive loop becomes a core part of the centaur's identity.

Agency in this context means more than being able to act. It means being able to form and revise commitments in light of evidence and value. A centaur is thus not just faster at producing documents than a human alone. It is, in the best case, better at tracing consequences, surfacing alternatives, and resisting the pull of cheap satisfaction. Moving from monkeys to centaurs is moving from scattered, uncoordinated interaction to joint agency with a recognizable trajectory.

9.2 tau(t) as living truth versus flat optimization objectives

Most current AI systems are trained and evaluated according to flat optimization objectives: loss functions, reward signals, engagement metrics, cost curves. These objectives can be highly detailed, but they do not carry an intrinsic sense of direction beyond "more of this" or "less of that." They are agnostic about whether the patterns they optimize correspond to anything like truth, justice, or wisdom.

logostics introduces tau(t) as an alternative picture. Instead of a fixed objective, tau(t) is a living manifold of truth that changes over time as more of reality is contacted and integrated. It is not simply "whatever the model predicts best" or "whatever maximizes reward." It is a trajectory of alignment between what agents believe and value and what is actually the case at multiple levels: physical, social, moral, spiritual.

For a centaur, tau(t) is not a parameter in an algorithm but a felt constraint. It shows up as the sense that some lines of inquiry deepen contact with reality while others drift into self-serving fantasy. It is present when the pair abandons a neat theory because new evidence does not fit, or when it refuses to deploy an effective rhetorical strategy because it would mislead. In this sense, tau(t) functions like a slope on the landscape of possible thoughts and actions: some directions lead toward greater coherence and integrity, others lead away.

The monkey regime, by contrast, orients primarily to flat objectives such as engagement, throughput, and local convenience. A centaur does not ignore these; it still needs attention, resources, and time. But it treats them as constraints to be managed, not as ultimate goals. The normative horizon of stage 3 is a world in which human–AI pairs use models to explore more of tau(t), not simply to better optimize surrogate metrics. This does not require mystical certainty about truth. It requires humility and a willingness to let evidence and conscience overrule the path of least resistance.

9.3 Compression with gain: fewer artifacts, deeper alignment

One of the puzzles of the AI-saturated world is how to measure progress when sheer volume of artifacts is cheap. If a monkey regime can produce millions of pages with minimal effort, an alternative criterion is needed. Compression with gain offers such a criterion. It asks: can a centaur produce fewer artifacts that nonetheless capture more structure, insight, and alignment than larger piles of unstructured output?

Compression here is not merely shorter text. It is the act of distilling a complex space of possibilities into a smaller set of representations that preserve and enhance what matters. Gain is the increase in clarity, predictive power, or ethical traction that results. A centaur that revisits its own corpus, pruning, refactoring, and integrating, is engaged in compression with gain. It is not satisfied with scattering countless drafts; it wants to produce statements, models, and designs that stand up under repeated use and scrutiny.

Generative AI can assist with this process if it is used deliberately. Models can help cluster themes, detect redundancies, and propose unifying formulations. They can simulate critical readers, propose counterexamples, and identify places where the argument is weak. But without a centaur's orientation to tau(t), these same capabilities simply produce more variations on the same noise.

In stage 3, compression with gain becomes a key performance measure. It changes what counts as success. Instead of celebrating an author for producing a thousand articles, we might look at whether their tenth synthesis makes the first nine obsolete by integrating and transcending them. Instead of measuring an organization by the number of AI-assisted reports it generates, we might ask whether its key decisions become more coherent and less reactive over time.

This criterion also constrains AI development. Models tuned for compression with gain would prioritize assisting in refinement, integration, and critique, rather than maximizing engagement. They would be assessed on how well they help centaurs reduce confusion, not on how much they can produce on demand.

9.4 Centaur biographies as a new genre of intellectual history

If centaurs become important units of agency, their lives will require new forms of narration. Traditional biographies focus on individual humans: their upbringing, influences, works, and legacy. Stage 3 calls for centaur biographies: accounts of human–AI pairs as evolving cognitive and ethical entities.

A centaur biography would track not only the human's education and experiences, but also the history of their interactions with specific models and systems. It would describe how their prompting style changed over time, how they responded to model failures, how their joint workflows evolved, and how their sense of responsibility deepened or frayed. It would treat chat logs, version histories, and internal documents as primary sources, alongside letters and interviews.

Such biographies would also foreground the dynamics of alignment. They would show how the centaur's relation to tau(t) shifted: moments when it compromised under pressure, moments when it reversed course in light of new evidence, moments when it resisted the seductions of engagement or speed. The central question would not be "What did this person produce?" but "How did this pair learn to see more truly and act more responsibly, or fail to do so?"

In this genre, failure becomes as instructive as success. A centaur that drifted into self-deception, or that allowed optimization for local metrics to corrode its commitments, teaches as much about the hazards of people–AI collaboration as a centaur that remained oriented to tau(t). Biographies could map patterns across cases, identifying recurring traps and protective practices.

Writing such histories would itself be a centaur task. Human historians, working with AI tools, would reconstruct trajectories from vast digital traces. They would need to decide what to anonymize, what to foreground, and how to respect privacy while extracting lessons. Done well, centaur biographies could serve as guides for future apprentices, much as lives of scientists, saints, and artists have traditionally served as guides for those entering their paths.

In imagining stage 3 as a horizon populated by logostic centaurs oriented to tau(t), this section does not claim that such a world will inevitably arrive. It claims only that without some version of this horizon, the monkey regime has no clear alternative. Monkeys can be made more efficient. Engagement curves can be flattened or steepened. But without a vision of joint agency aligned with living truth, the people–AI story risks becoming nothing more than a tale of ever more sophisticated typewriters in ever larger farms. The centaur, by contrast, offers a way to ask a different question: not just what can be generated, but who we become by generating it together.

10. Dynamics and Attractors: Why 2025 Stayed in Stage 1

Stage 1 was not just an awkward transitional phase on the way to something better. It was a genuine attractor: a configuration of technology, economics, psychology, and politics that actively held societies in place. Even as proto-apprenticeships and centaur-like dyads began to appear, the large-scale dynamics kept pulling behavior back toward the monkey regime. Understanding these dynamics is essential for anyone who wants to design realistic exits from Stage 1 rather than assuming that time and familiarity will automatically deliver more mature people–AI relations.

This section traces four of the strongest forces that kept 2025 anchored in Stage 1. Structural incentives around advertising, engagement, and quarterly earnings rewarded volume and virality over depth and care. Cognitive laziness made the comfort of stochastic fluency appealing, even to those who suspected they were outsourcing too much. Political polarization created powerful incentives to weaponize AI output for rhetorical advantage. Finally, path dependence ensured that early patterns of cheap text and weak filtering would shape future trajectories long after their initial causes faded.

10.1 Structural incentives: advertising, engagement, quarterly earnings

At the largest scale, the behavior of platforms and major AI providers was governed by a simple fact: they existed inside an economic system that measured success in terms of revenue, growth, and shareholder value. For most, the dominant revenue model involved some combination of advertising, subscriptions, and enterprise contracts. Each of these, in different ways, favored rapid deployment of generative capabilities and high engagement with AI-driven interfaces.

Advertising revenue depended on attention. The more time users spent inside an ecosystem, the more ads could be served, the more data could be collected, and the more precisely ads could be targeted. Generative tools that increased engagement—by making search feel like conversation, by turning static documents into interactive sessions, or by enabling endless content generation—were naturally favored. There were no direct financial rewards for reducing the number of low-quality artifacts or for helping users generate less but better. The metric that mattered was usage, not compression with gain.

Engagement itself became both means and end. Product teams optimized interfaces to encourage more prompts, more clicks, more follow-up questions. AI assistants were integrated into email clients, office suites, and messaging apps with the explicit goal of making them “stickier.” When usage metrics rose after AI features were introduced, this was taken as evidence of success. Whether the resulting outputs improved understanding, strengthened

relationships, or contributed to better decisions was harder to measure and, therefore, easy to defer.

Quarterly earnings added temporal pressure. Publicly traded companies had limited patience for slow, deliberative transitions to Stage 2 practices that might initially reduce output or require retraining. The demand for visible AI "wins" within short time horizons encouraged shipping features quickly, marketing them aggressively, and worrying about long-term epistemic consequences later. Safety and alignment teams existed, but they were often framed as cost centers, not revenue drivers.

These structural incentives acted like gravity, pulling implementations toward the monkey regime. Even executives who personally worried about information overload or manipulation faced a system that rewarded them for delivering more features, more engagement, and more monetizable interactions. Without changes to these underlying incentive structures, appeals to ethics or calls for "responsible AI" had limited leverage.

10.2 Cognitive laziness and the comfort of stochastic fluency

At the psychological level, Stage 1 was stabilized by a simple human tendency: people prefer easier ways of doing difficult things, especially when the costs of the shortcut are not immediately visible. Generative AI offered a particularly seductive form of ease: the ability to turn vague intentions into fluent text without passing through the discomfort of drafting, rephrasing, and wrestling with one's own confusion.

Stochastic fluency—the model's capacity to generate smooth, confident language in almost any register—was especially comforting. It covered hesitations, filled gaps, and made the user sound more knowledgeable and composed than they might feel. For someone facing an inbox full of emails, a stack of assignments, or a looming deadline, the relief of watching the AI "handle it" was real. The brain, already taxed by modern life, gratefully accepted the reduction in immediate effort.

The laziness here was not moral failure so much as cognitive economizing. People are wired to conserve mental energy when possible. Before generative AI, certain tasks—long emails, careful explanations, structured arguments—were sufficiently effortful that doing them at all signaled some level of commitment. In 2025, the same tasks could be completed with much less investment. It was natural to ask: why strain when the model can do it?

What made this tendency stabilizing was the time lag between shortcut and consequence. The erosion of felt authorship, the weakening of argument-building muscles, the dependency on external pattern generators—all these emerge slowly. The feedback is diffuse. A user who relies on AI to draft emails does not receive immediate, clear negative signals that their own writing skills are degrading. On the contrary, they often receive positive signals: praise for clarity, faster

responses, fewer misunderstandings. The long-term cost is abstract, the short-term benefit concrete.

Cognitive laziness also manifested as reluctance to verify. Checking model outputs against primary sources, running independent calculations, or challenging plausible answers all require deliberate effort. When the first answer feels "good enough," it is psychologically easy to stop there, especially in low-stakes settings. Over time, this habit generalizes. Even in higher-stakes contexts, the baseline of scrutiny may be lower than it should be because users have become habituated to accepting fluent output as adequate.

In Stage 1, then, generative AI and cognitive laziness formed a mutually reinforcing loop. The more people experienced effortless fluency, the less they tolerated the discomfort of cognitive struggle. The less they struggled, the more they needed AI to cover their discomfort. Breaking this loop in the direction of Stage 2 requires not only institutional redesign but also a revaluation of effort: treating certain kinds of mental exertion as intrinsically valuable, not as inefficiencies to be eliminated.

10.3 Political polarization and the weaponization of AI output

Politics in 2025 was already polarized before AI arrived. Generative tools did not create that polarization, but they offered new ways to weaponize it. Parties, movements, and factions discovered that AI could mass-produce messages tailored to specific audiences, generate plausible-sounding arguments for any position, and fabricate or embellish narratives that fit pre-existing grievances. The monkey regime was particularly attractive in this domain because it rewarded speed, volume, and emotional impact.

The logic of polarized politics aligns well with the logic of generative tools. In a conflict-driven environment, the goal is often not to approach tau(t) but to win information battles: to mobilize supporters, demoralize opponents, and shape the perceptions of the undecided. AI-generated memes, talking points, and "explainer threads" can be customized for different segments, rapidly tested for effectiveness, and iterated in response to real-time analytics. The cost of inventing new frames, slogans, or attacks drops dramatically.

This dynamic further destabilized incentives for careful use of AI. Actors who chose to adhere to norms of accuracy and restraint risked losing ground to those willing to exploit AI for more aggressive tactics. If one side floods local channels with AI-generated narratives that play on fear and outrage, the other side faces pressure to respond in kind or risk being drowned out. Calls for higher standards thus collided with the perceived necessity of "not unilaterally disarming" in an arms race of synthetic rhetoric.

The weaponization of AI output also complicated public perception of generative tools. For some, AI became associated primarily with deepfakes, disinformation, and manipulative

persuasion. For others, it was seen as an indispensable instrument for defending one's side in a hostile environment. In neither case was the dominant frame one of patient apprenticeship or centaur-like exploration of shared reality. The battlefield conditions of online politics made Stage 1 behaviors not only tempting but, from the perspective of partisan actors, rational.

Under such conditions, proposals for more deliberate, cross-partisan uses of AI—such as jointly exploring policy trade-offs, generating balanced briefs, or testing arguments for fairness—struggled to gain traction. The short-term tactical advantages of weaponized generative output overshadowed the long-term strategic advantages of a more coherent, less inflammatory public sphere. As long as polarization remained high and incentives rewarded conflict, the monkey regime retained a strong hold over political uses of AI.

10.4 Path dependence: why early cheap text locks in future trajectories

Finally, Stage 1 persisted because early choices about how to deploy generative AI created path dependencies that were hard to reverse. Once a critical mass of cheap text, images, and code had entered the world, subsequent decisions had to be made in the context of that accumulated output and the expectations it created.

On the supply side, organizations that invested heavily in AI-driven content pipelines built workflows, tools, and metrics around high-volume production. Marketing departments staffed and trained for continuous generation of AI-assisted campaigns. News aggregators and content farms optimized their infrastructure for ingesting and lightly editing AI drafts. Rebuilding these systems around Stage 2 principles—slower production, more editing, deeper verification—would have required not only technical changes but also cultural and financial ones.

On the demand side, audiences habituated to certain patterns. Users grew accustomed to instant answers, personalized summaries, and infinite scrolls of new material. Attention spans and reading habits adjusted accordingly. The idea of waiting for fewer, more carefully curated texts began to feel alien or elitist. Educational systems that had allowed assignments to be completed with AI assistance found it increasingly difficult to revert to modes of evaluation that required sustained, unaided writing or problem-solving.

There was also a data feedback loop. Early AI-generated content entered training sets for future models, either directly or through its influence on online corpora. This meant that the stylistic quirks, biases, and shallow patterns of Stage 1 output risked being amplified in subsequent generations of models. Without deliberate filtering to exclude or down-weight such material, the models themselves would become more aligned with the monkey regime, making it harder to use them as instruments for deeper inquiry.

Path dependence showed up in regulation as well. Early rules, norms, and legal precedents set expectations about what was considered acceptable or normal AI use. Once courts had ruled on

certain cases, once agencies had written guidelines, and once industry standards had emerged, these frameworks constrained future options. Shifting from a world that tolerated or even encouraged AI-saturated communication to one that demanded stricter editorial control would involve not just new policies but the renegotiation of existing agreements and vested interests.

In short, 2025's Stage 1 patterns were not just a temporary mood; they were the foundation of an emerging socio-technical infrastructure. Trajectories laid down in this period—about what counts as productivity, how quickly content should appear, how much verification is expected, and what users demand—will continue to shape possibilities for years to come. Recognizing this path dependence is not a counsel of despair, but a reminder that moving toward Stage 2 and Stage 3 will require deliberate interventions strong enough to bend trajectories already in motion, not just incremental adjustments at the margins.

Together, these dynamics—structural incentives, cognitive tendencies, polarized politics, and path-dependent infrastructures—explain why 2025 remained largely in Stage 1 despite the presence of better possibilities. They also indicate where leverage might lie: in redesigning business models that reward depth, fostering cultures that value cognitive effort, de-escalating political arms races, and intervening early in training pipelines and institutional norms. Without such efforts, Stage 1 will not simply fade; it will harden into the default operating system of people–AI civilization.

11. Design Principles for Exiting the Monkey Regime

If Stage 1 is an attractor, exiting it will not happen by accident. It will require deliberate counter-design: choices that run against the grain of existing incentives, habits, and infrastructures. The goal is not to abolish generative AI or to return to a pre-digital scarcity of text. It is to build structures in which people–AI collaboration tends toward apprenticeship and centaur forms of agency rather than toward unstructured, responsibility-light generation.

This section sketches four design principles for that transition. First, reintroducing carefully placed friction can slow down generation where it matters and speed up review where it is currently neglected. Second, visible apprenticeship paths can make AI-intensive craft a recognized domain of skill rather than an invisible improvisation. Third, rewiring metrics can shift attention from short-term clicks to long-horizon coherence and care. Finally, institutional experiments—labs, scriptoria, and logostic communities—can serve as testbeds for new norms and practices that might later spread.

11.1 Reintroducing friction: slowing down generation and speeding up review

One of the defining features of the monkey regime is that generation is nearly frictionless while review and reflection are costly and optional. Design for Stage 2 and Stage 3 reverses that asymmetry wherever stakes are high. Instead of asking, “How can we make it even easier to generate?” the guiding question becomes, “Where should it be harder to publish and easier to think?”

Reintroducing friction does not mean inserting pointless delays everywhere. It means selectively adding resistance at thresholds where words and outputs acquire real consequence. For instance, internal tools might allow rapid generation of drafts but require extra steps for anything that will be sent outside the organization: a mandatory checklist, a review by a named editor, or a time delay that encourages a second look. Public-facing systems might distinguish clearly between “scratchpad” and “signature” modes, with different levels of scrutiny and logging.

On the review side, friction should be reduced. It should be simpler to annotate, critique, and compare AI-assisted outputs than to produce them. Interfaces can foreground side-by-side views, version histories, and built-in critique prompts that encourage users to ask models to challenge themselves. Workflows can allocate explicit time for review, rather than treating it as something tacked on after the “real work” of generation.

Over time, these design choices can shift habits. When people expect that anything important they generate will encounter resistance—questioning, scrutiny, revision—they are more likely to anticipate that resistance and to generate with review in mind. When they know that reviewing and challenging outputs is supported and rewarded, they are more likely to invest effort there. Friction thus becomes a form of pedagogy: it teaches users that not all clicks are equal, and that some should carry the weight of responsibility.

11.2 Building visible apprenticeship paths for AI-intensive work

In 2025, much of the craft of working well with AI was invisible. Individuals who developed good practices did so informally, and their knowledge rarely translated into shared standards. Exiting the monkey regime requires making these crafts visible and building explicit paths for others to learn them.

A visible apprenticeship path has at least three components. First, it has named roles. Instead of every knowledge worker being assumed to be “fine with AI,” organizations recognize roles like AI editor, AI workflow designer, or centaur practitioner. These roles come with expectations, time allocations, and authority to say no when proposed uses of AI are unsafe or misaligned.

Second, it has structured stages. Apprenticeship is not a binary of “untrained” versus “expert.” It involves progression. Training programs can mark transitions from novice to intermediate to advanced user with identifiable milestones: demonstrating competence in prompt design, error detection, verification strategies, and ethical judgment in AI-intensive tasks. These milestones can be assessed using realistic scenarios rather than simple quizzes.

Third, it has public artifacts. Apprenticeship paths become visible when bodies of example work, annotated workflows, and postmortems of failures are shared within and across institutions. Rather than hiding AI use as a competitive secret or a source of shame, organizations can publish case studies: where AI helped, where it misled, and what was learned. This creates a culture in which learning how to work with models well is a collective project, not an individual struggle.

When apprenticeship paths are visible, people can aspire to them and be recognized for pursuing them. The monkey regime, which treats AI use as a generic, unskilled activity, gives way to a landscape in which some people are known as careful stewards of human–AI collaboration. That recognition is essential if we want responsibility and craft, not just enthusiasm and improvisation, to define the next phase.

11.3 Rewiring metrics: from clicks to long-horizon coherence and care

Metrics are the nervous system of modern institutions. They tell organizations what to pay attention to and what to ignore. In Stage 1, the dominant metrics—clicks, impressions, time-on-site, task completion counts—rewarded behaviors that generate more activity, not necessarily better outcomes. To exit the monkey regime, metrics must be rewired to track something closer to tau(t): long-horizon coherence and care.

This rewiring operates at multiple levels. For platforms and media outlets, it might mean measuring trust and understanding rather than just engagement. Surveys, longitudinal studies, and controlled experiments can estimate whether users come away from AI-mediated interactions better informed, more confused, or more polarized. Content that consistently improves understanding can be promoted; content that merely provokes anger or addiction can be demoted, even if it performs well on raw engagement.

For organizations that rely on AI in internal workflows, metrics can track downstream effects rather than just immediate throughput. Instead of celebrating how many AI-assisted documents are produced, they can ask how often those documents need correction, how many errors slip through, and how decisions made on their basis fare in the real world. Metrics of decision quality, error rates, and alignment with stated values can be fed back into performance evaluations and process redesign.

For individuals and centaurs, metrics shift from quantity to trajectory. A researcher might track not just the number of papers produced with AI assistance, but whether successive works resolve earlier confusions or simply proliferate parallel lines of thought. A policy team might ask whether their recommendations become more consistent and less reactive over time as they use AI, or whether they are being tugged by whatever the latest model output suggests.

These new metrics will be fuzzier and harder to optimize than clicks. That is precisely the point. They align incentives with the difficult, slow work of approaching tau(t). They encourage institutions to tolerate some reduction in immediate output in exchange for more robust, less self-destructive patterns over years and decades. Without such rewiring, even the best-designed apprenticeship programs risk being undermined by background pressures to produce more, faster, and louder.

11.4 Institutional experiments: labs, scriptoria, and logostic communities

Finally, exits from Stage 1 will not emerge from theory alone. They will require institutional experiments: bounded spaces where different rules are tried, refined, and, if they work, exported. Three archetypes—labs, scriptoria, and logostic communities—illustrate how such experiments might be structured.

Labs are environments explicitly dedicated to exploration and failure. An AI lab in this sense is not only a technical research group but a socio-technical workshop: a place where people test new workflows, prompts, verification practices, and governance mechanisms. Labs can run controlled trials of alternative designs: different friction patterns, different review protocols, different ways of representing uncertainty. Their findings can inform policy and practice beyond their walls.

Scriptoria, in this context, are institutions that commit to high standards for a specific class of outputs: scientific articles, legal rulings, medical guidelines, curricula, or liturgical texts. They can adopt strong editorial norms: every AI-assisted contribution must pass through named human editors; all AI involvement must be documented; decisions about publication must explicitly consider alignment with stated values. Scriptoria function as exemplars: they show that it is possible to use AI intensively without surrendering to the monkey regime.

Logostic communities are groups that explicitly orient their people–AI practices toward tau(t). They might be research teams, congregations, activist circles, or interdisciplinary institutes. What defines them is a shared commitment to treating AI as an instrument for contacting and responding to living truth, not merely as a tool for efficiency or persuasion. Such communities might adopt internal covenants about how and where AI is used, how disagreements are worked through, and how successes and failures are interpreted. They are places where centaur biographies can be written in real time.

These experiments need not be grandiose. They can be small, local, and provisional. What matters is that they exist as alternatives to the default monkey farms: places where people and models are invited into more demanding relationships. Over time, successful experiments can inform standards, regulations, and expectations elsewhere. Failed experiments can map dangers and dead ends. In both cases, the act of experimenting is itself a way of resisting the passivity that defines Stage 1.

Taken together, these design principles—friction, apprenticeship, metric rewiring, and institutional experimentation—do not guarantee a smooth transition to Stage 2 or Stage 3. They do, however, mark where leverage is likely to be found. They suggest that exiting the monkey regime is less about discovering a new technology and more about choosing a different civilization game: one in which we measure ourselves not by how many typewritten pages we can produce, but by how deeply we and our machines together can move toward what is true, good, and worth keeping.

12. A Research Agenda: Measuring Monkeys, Apprentices, and Centaurs

If "monkeys, apprentices, and centaurs" are to be more than evocative metaphors, they need to be operationalized. Otherwise, Stage 1 remains a moral diagnosis rather than a subject of empirical inquiry, and Stage 2 and Stage 3 remain aspirations rather than design targets. A research agenda for people–AI co-evolution must therefore answer a deceptively simple question: how would we know, in practice, whether a given person, team, or institution is mostly behaving like a monkey, an apprentice, or a centaur?

This section sketches such an agenda. It begins by proposing operational indicators of Stage-1 monkey behavior at micro and meso scales. It then suggests ways to detect emergent scriptorium practices in organizations. Next, it outlines methods for studying centaur pairs as primary units of analysis rather than anomalies. Finally, it argues for longitudinal studies that track people–AI co-evolution over years, not weeks, so that transitions among stages and their consequences can be observed rather than merely inferred.

12.1 Operational indicators of stage-1 monkey behavior

To study the monkey regime empirically, we need observable patterns that signal "monkey-ness" without requiring speculative psychological access. Several classes of indicators are promising.

One class concerns **ratio and timing of generation versus revision**. In Stage 1, users often generate long outputs in a few keystrokes and spend minimal time editing or verifying. Log data

can capture the frequency of “generate” actions, the average length of outputs, and the time spent between generation and submission. A high ratio of tokens generated to tokens edited, combined with short review intervals, is a plausible signature of monkey behavior, especially in tasks that would normally demand careful thought.

A second class involves **prompt and interaction structure**. Monkey usage tends to show shallow prompt patterns: single-shot requests, vague instructions, and little decomposition of complex tasks. By contrast, even proto-apprenticeship tends to involve multi-step dialogues: asking for outlines, then examples, then critiques. Quantitatively, this suggests metrics such as average conversation depth per task, proportion of sessions that include explicit requests for critique or verification, and frequency of prompts that specify constraints related to truth, uncertainty, or ethics.

A third class focuses on **verification and cross-checking behavior**. Monkeys rarely ask the model to justify claims, list sources, or highlight uncertainties. They seldom cross-check outputs against independent tools or datasets. Observable indicators here include how often users request alternative answers, ask “Are you sure?” style questions, or bring in external references. A low rate of such behaviors, particularly in high-stakes domains, suggests Stage-1 patterns.

Finally, there are **downstream error and correction patterns**. In contexts where AI-assisted outputs are later found to be wrong or harmful, researchers can analyze whether errors stemmed from predictable model weaknesses combined with absent review. A recurring pattern of uncorrected, easily detectable errors slipping through is another marker of monkey-like use: generation without sufficient responsibility.

Taken together, these indicators do not pathologize individual users. They provide a way to say, with some precision, “this environment is rewarding Stage-1 behavior,” and to quantify whether interventions are nudging practice toward apprenticeship or centaur-like patterns.

12.2 Detecting emergent scriptorium practices in organizations

If Stage 2 is a scriptorium of apprentices, we should be able to detect its early emergence in organizations before it becomes official doctrine. The research task is to identify **meso-level signatures** of scriptorium-like behavior.

One signature is the **presence of explicit AI usage norms in local teams**, even when absent from corporate policy. Ethnographic and survey work can look for teams that have written or unwritten rules such as: “We never send AI-generated content externally without human review,” “We always document when AI was involved,” or “We use AI for first drafts but never

for final recommendations." The existence and diffusion of such norms are strong indicators that Stage-2 craft is forming.

Another signature is **institutional investment in review and editing roles** specifically for AI-intensive work. This can be measured by tracking whether organizations create roles like "AI editor," "AI workflow lead," or "model safety liaison," and whether these roles carry decision-making power or are merely symbolic. Budget allocations, job descriptions, and promotion criteria can all be studied to see whether organizations are making editorial capacity a first-class concern.

A third signature is **the structure of AI-related training and support**. Are training sessions limited to showing where the "magic button" is, or do they include modules on model limitations, verification strategies, and ethical use? Do internal knowledge bases contain examples of good and bad AI-assisted work, annotated with commentary? The richness and reflexivity of training materials can be coded and compared across institutions.

Finally, scriptorium practices show up in **the topology of organizational workflows**. In Stage 1, workflows often terminate as soon as an AI output is generated and superficially approved. In Stage 2, new steps appear: explicit review gates, mandatory critical passes, and periodic audits of AI-assisted projects. Process mining techniques can detect whether organizations have inserted such gates and how often they are actually used.

By identifying and tracking these features, researchers can map where scriptoria are emerging, how they differ across sectors, and which conditions support their survival and growth in environments still dominated by monkey farms.

12.3 Studying centaur pairs as units of analysis

Centaur-like dyads—stable human–AI partnerships—are arguably the most interesting entities in the people–AI ecosystem, yet they are the least studied. A research agenda for Stage 3 must treat these pairs as primary units of analysis, not anomalies.

Methodologically, this suggests **multi-modal case studies** that combine log analysis, interviews, and artifact review. Researchers can recruit centaur-like practitioners—people who have long-running relationships with specific models and workflows—and, with consent, analyze their interaction history over months or years. They can examine patterns such as how tasks are decomposed, how often the human challenges the model, and how the pair's practices change in response to errors or new capabilities.

Key variables include **division of cognitive labor**, **reflexive dialogue**, and **alignment behavior**. How much of the work is the human doing in conceptual framing versus detail elaboration?

How often does the human ask the AI to critique its own outputs or simulate opposing views? Are there visible moments where the centaur chooses a path that is less efficient in the short term but better aligned with stated values or long-term goals?

Researchers can also look at **performance trajectories**. Does the centaur's output, over time, show signs of compression with gain—fewer but more integrated artifacts, clearer lines of argument, fewer repeated mistakes? Comparing centaur pairs to comparable humans working without AI can reveal not only performance differences but differences in how knowledge is structured and updated.

Ethically, studying centaurs requires strong safeguards: anonymization, clear boundaries on what can be analyzed, and respect for the fact that chat logs and work products may contain intimate traces of thought. But if done carefully, this line of research could yield a new kind of cognitive ethnography: not of individuals alone, but of hybrid agents whose identities and responsibilities straddle the human–machine boundary.

12.4 Longitudinal studies of people–AI co-evolution

Finally, because Stage transitions are temporal, any serious research agenda must be longitudinal. Cross-sectional snapshots can show where things stand; they cannot show how monkeys become apprentices or apprentices become centaurs—or how they regress.

Longitudinal studies can take several forms. One is **panel studies of users** who consent to have their AI interactions sampled over time, alongside surveys and performance measures. Researchers can observe how their prompting styles, verification habits, and felt authorship change as they gain experience or as models evolve. They can relate these changes to outcomes such as skill development, job performance, and psychological well-being.

Another form is **organizational longitudinal studies**. By embedding researchers within institutions undergoing AI integration, it becomes possible to track how policies, workflows, and cultures change across months and years. Do early bans give way to structured adoption? Do informal scriptoria get formalized or suffocated? How do error incidents or public scandals reshape internal practices?

A third form is **cohort-based experimental interventions**. Groups of users or teams can be given different training, tools, or metrics aligned with Stage-2 or Stage-3 principles, and their trajectories compared to control groups operating under standard Stage-1 conditions. Outcomes can include not only productivity and error rates but measures of epistemic humility, resilience to misinformation, and capacity for coherent long-term projects.

Longitudinal designs are also necessary to study **feedback from AI-generated content into future models and practices**. As synthetic text becomes training data, and as users adapt their behavior to what models make easy or difficult, a co-evolutionary loop forms. Tracking this loop requires connecting data from multiple generations of models, interaction logs, and cultural outputs over extended periods.

In all these designs, the central aim is the same: to treat people–AI relations as evolving systems with identifiable dynamics, not as static snapshots of "use." Only then can we answer the central questions of this paper empirically: How sticky is the monkey regime? Under what conditions do proto-apprenticeships grow into scriptoria? What distinguishes centaurs that deepen their alignment with tau(t) from those that drift? And, most practically, which interventions reliably bend trajectories away from cheap stochastic fluency and toward fewer, truer words shared between humans and their machines?

13. Conclusion: First Drafts, Not Final Judgments

The story of 2025 is not the story of what people and AI will be forever. It is a noisy, uneven, often embarrassing first draft: a year when billions of people suddenly acquired access to systems that could flood the world with plausible language, and most of them did the obvious thing. They tried it, they played with it, they used it to cut corners, and they improvised norms as they went. The monkey regime is the predictable result of that improvisation under existing incentives. It looks chaotic from one angle and strangely coherent from another, because the same structural forces show up everywhere: speed over depth, engagement over truth, convenience over responsibility.

This paper has treated 2025 not as a final verdict on people–AI civilization but as an early chapter from which multiple futures can unfold. The three-stage framework—monkeys, apprentices, centaurs—is offered less as a rigid classification and more as a vocabulary for noticing patterns and designing alternatives. Stage 1 describes where we mostly were. Stage 2 and Stage 3 describe where we mostly were not, but might still choose to go.

13.1 2025 as a noisy first draft of people–AI civilization

Seen in historical perspective, 2025 resembles other moments when a powerful new medium became cheap and ubiquitous before cultures learned how to handle it. Early printing presses produced not only scripture and science but also pamphlets, propaganda, and dubious almanacs. Early broadcast media brought both public service announcements and demagogues into millions of homes. When the web appeared, spam and pop-up ads were not bugs; they were the first stable business models that fit the new terrain. In each case, the first draft of

usage overexpressed the medium's capacity for noise and underexpressed its capacity for shared understanding.

Generative AI in 2025 fit this pattern. Cheap text led to more text. Cheap images led to more images. Cheap code led to more code. Most of these artifacts were small, local, and quickly forgotten: school assignments written in a rush, emails polished just enough to pass, sermons padded, spam multiplied. Some artifacts were more ambitious—research papers, policy briefs, creative works—but they too were shaped by the default affordances of the monkey regime: rapid drafting, limited verification, and often thin reflection on what it meant to sign one's name to something co-written by a machine.

Calling this a first draft is not an excuse. It is a way of naming incompleteness. First drafts matter. They establish habits, set expectations, and populate the training data of future models. But they are still drafts. They can be annotated, corrected, and radically revised. The question is not whether 2025 was messy; it was. The question is whether later years will treat that mess as something to be cleaned up or as the permanent baseline for people–AI interaction.

13.2 The dangers of staying in the monkey regime too long

The monkey regime is tolerable in short bursts. It is not tolerable as a permanent operating system. Staying in Stage 1 too long carries at least three kinds of danger: cognitive, institutional, and moral.

Cognitively, prolonged dependence on stochastic fluency risks hollowing out skills that took centuries to build: careful argumentation, slow reading, precise writing, and the ability to sit with confusion rather than papering it over with smooth language. If every hard thought can be instantly wrapped in plausible prose, it becomes easier to mistake articulation for understanding. Over time, this can produce a culture of superficially competent texts backed by increasingly shallow grasp of what those texts claim.

Institutionally, a long-term monkey regime corrodes trust. When organizations flood their stakeholders with AI-assisted communications that are polished but weakly grounded, people eventually notice. They may not always articulate what feels off, but they sense that something has changed in the relationship between voice and responsibility. Policies, public statements, and even personal letters begin to feel less like expressions of considered judgment and more like products of a machine-mediated PR layer. Once that perception sets in, efforts to reintroduce sincerity and care face uphill battles.

Morally, the danger is drift away from any living sense of tau(t). If generative systems are used primarily to optimize for engagement, internal metrics, or partisan advantage, they will gradually reshape what people treat as real and worth caring about. A world saturated with monkey-level output is one where genuine signals struggle to stand out, where warnings about

genuine risks and invitations to genuine hope compete with an endless churn of attention-capturing noise. In such a world, it becomes harder to hear conscience, evidence, or the quiet demands of those who are not algorithmically loud.

The longer Stage 1 persists as the default, the more these dangers become background conditions rather than visible problems. The point is not that everyone will become shallow, cynical, or confused. The point is that the path of least resistance will tilt in that direction, and that climbing toward more demanding practices will feel increasingly like swimming against the current.

13.3 The promise of apprenticeship and logostic centaurs

Against this backdrop, the promise of apprenticeship and logostic centaurs is modest and radical at once. Modest, because it does not claim that better people–AI relations will solve all human problems. Radical, because it insists that how we structure those relations will either deepen or erode our capacity to respond to what is real.

An apprenticeship regime promises to reframe AI not as a replacement for thought but as a context for learning to think differently and better. In such a regime, novices are not given models and left to fend for themselves. They are guided into practices that emphasize decomposing problems, probing model limits, cross-checking claims, and taking responsibility for final outputs. They learn that the point of using AI is not to avoid difficulty but to face more meaningful forms of difficulty: the kind that arises when one's own assumptions are challenged, when contradictory evidence appears, or when the easy rhetorical win is misaligned with deeper commitments.

Logostic centaurs extend this promise. They sketch a way of life in which a human and a model together become capable of sustained trajectories toward tau(t). In that horizon, the model's pattern-recognition and generative power are not ends in themselves. They are instruments for contacting, testing, and articulating truths that neither partner could reach alone. A centaur that is attentive to tau(t) does not simply generate more content; it systematically reduces confusion, clarifies stakes, and resists temptations to optimize for something flatter than living truth.

The promise is not utopian. Centaurs can fail, drift, and be corrupted. Apprentices can internalize bad habits. Scriptoria can ossify into complacent guilds. But the existence of these modes of failure presupposes something worth failing at: a practice of people–AI collaboration that aspires to more than speed, novelty, or dominance. That aspiration, once articulated, is itself a resource. It gives designers, educators, and practitioners something to point to when they say, "We could do this differently."

13.4 Final reflections on authorship, responsibility, and the texts that will outlive us

Generative AI unsettles intuitions about authorship at a basic level. Who wrote this sentence? The person who typed the prompt, the model that produced the draft, the editor who revised it, or the institutions and communities whose norms shaped the whole process? In Stage 1, the practical answer was often "no one in particular." The text appeared; it was useful enough; everyone moved on. Responsibility dissolved into the interface.

Exiting the monkey regime requires recovering a thicker sense of authorship. Not in the romantic sense of solitary genius, but in the more ancient sense of being answerable for what one sets in motion. When a human signs a document, publishes a post, or endorses a model-assisted recommendation, they are authoring consequences. They may have delegated much of the drafting to a machine, but they have not delegated accountability. Apprenticeship and centaur practices both depend on making that accountability visible, teachable, and livable.

The texts that will outlive us—the ones future historians, centaur biographers, or logostic communities will actually care about—are unlikely to be the mass-produced outputs of the monkey regime. They will be the artifacts in which people and models together confronted hard questions without flinching, integrated disparate domains of knowledge, and chose integrity over expedience. They may be fewer in number than the flood of everyday AI-generated text, but they will carry more energy per word.

This paper is itself only a first draft in that direction. It has used the very tools it analyzes. It has relied on the fluency of a model to articulate concerns about fluency, and it has tried, within that paradox, to point toward more demanding uses of the same capacities. Whether it succeeds will not be known quickly. The real test will be whether concepts like monkey regimes, scriptoria, and logostic centaurs help actual humans, in actual institutions, make different choices.

In the meantime, each prompt and each acceptance of an AI-generated sentence is a small rehearsal of the larger questions. What am I trying to do here? Am I merely filling a space, meeting a deadline, winning a point? Or am I, however clumsily, trying to move a little closer to tau(t)—to something truer, more coherent, and more worthy of being left behind? The answers will not be given by the model. They will be written in our own willingness to treat our words, and our machines' words, as more than noise in a flood.

14. References

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv preprint.

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. Quarterly Journal of Economics, forthcoming.

Carr, N. (2010). The shallows: What the Internet is doing to our brains. W. W. Norton.

Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7–19.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138.

Kasparov, G. (2017). Deep thinking: Where machine intelligence ends and human creativity begins. PublicAffairs.

Kalluri, P. (2020). Don’t ask if AI is good or fair, ask how it shifts power. Nature, 583, 169.

Lanier, J. (2010). You are not a gadget. Alfred A. Knopf.

McKinsey Global Institute. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.

OpenAI. (2023). GPT-4 technical report. arXiv preprint.

O’Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin Press.

Pew Research Center. (2025). How Americans view AI and its impact on people and society. Pew Research Center.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3–4), 379–423, 623–656.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), 495–504.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science, 333(6043), 776–778.

Tufekci, Z. (2017). Twitter and tear gas: The power and fragility of networked protest. Yale University Press.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., et al. (2022). Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.

Wheeler, J. A. (1990). Information, physics, quantum: The search for links. In W. H. Zurek (Ed.), Complexity, entropy and the physics of information. Addison-Wesley.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.

The author has published over 4,600 AI-assisted papers at: https://www.researchgate.net/profile/Douglas-Youvan/research . Google Scholar has indexed these preprints, and if you want a particular reference, the AI mode of a Google search (Gemini) has efficiently catalogued these preprints.