

KITALENT RESEARCH PAPERS

ESSAY · 2026

ARTIFICIAL INTELLIGENCE *Doesn't* EXIST

*Why What We Call AI Is
Actually Artificial Fluency*

THE MISNAMING, THE CATEGORY MISTAKE, AND THE
ONTOLOGICAL BOUNDARY BETWEEN SYNTAX AND JUDGMENT

ALESSIO MONTARULI

KITALENT ♦ RESEARCH

KITALENT RESEARCH PAPERS

Artificial Intelligence Doesn't Exist

*Why What We Call AI Is
Actually Artificial Fluency*

The Misnaming, the Category Mistake, and the
Ontological Boundary Between Syntax and Judgment

ALESSIO MONTARULI

COPYRIGHT

Copyright © 2026 Alessio Montaruli.
All rights reserved.

Artificial Intelligence Doesn't Exist: Why What We Call AI Is Actually Artificial Fluency

KiTalent Research Papers · 2026

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

To view a copy of this license, visit
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to share (copy and redistribute) the material in any medium or format under the following terms: Attribution (you must give appropriate credit and indicate if changes were made); NonCommercial (you may not use the material for commercial purposes); NoDerivatives (you may not distribute the modified material).

United States Copyright Office Registration:
Case #: 1-15168270217

This compact book edition was prepared from the author's Markdown paper manuscript.

CONTENTS

1. The Detachment of the Name from the Phenomenon	1
2. What LLMs Actually Do	8
3. Fluency Without Experience	19
4. The Political Economy of the Misnaming	27
5. Empirical Harms in Judgment-Class Domains	37
6. Adopting "Artificial Fluency"	50
Notes	57

1. The Detachment of the Name from the Phenomenon

The term “artificial intelligence” is applied to a wide range of contemporary software systems. The same term covers products that generate paragraphs, rank candidates, summarize clinical notes, answer legal questions, and draft reports. The term implies cognitive capacity. This paper examines whether that implication is supported by the systems it now describes.

The argument is not that the systems do not exist. They exist, operate at scale, and are economically significant. The argument is that the inherited term misdescribes them. The misdescription has a historical source. The term was introduced as the name of a research aspiration, not as a description of an existing technology.

The phrase acquired its canonical status in the proposal for the Dartmouth Summer Research Project of 1956. John McCarthy selected the phrase while attempting to name a new field distinct from cybernetics and related vocabularies. In the 1955 proposal, McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon described the project in a single sentence that defined the field's original ambition. The proposal stated that the study would proceed "on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."¹

That sentence specifies the content the term carried at its origin. "Artificial intelligence" did not refer to software capable of producing convincing language. It referred to a research program organized around the possible mechanization of intelligence itself. Learning, abstraction, language, originality, and self-improvement were treated as features that might be made precise enough for machine simulation. The proposal's confidence may now appear excessive, but that excess clarifies what the term meant. The term named a research demand, not a product category.

The research streams proposed for Dartmouth support this reading. Shannon approached cognition through information theory and brain models. Minsky proposed research on sensory and motor abstractions through which a machine might build an internal model of its environment. Rochester treated originality as guided randomness within constraints. McCarthy pursued formal languages for conjecture and self-reference. These projects differed in method, but they shared a working assumption: intelligence would require a structural relation between computation, representation, reasoning, and world. Surface imitation was not the target.²

The Dartmouth founders did not share a single theory of mind. They gathered several methods and assumptions under a new title. The relevant point is that the title referred to cognition as a problem. Even where the proposed methods were speculative, the object was learning, reasoning, abstraction, and possible self-improvement as machine functions. The term “intelligence” named an unfinished research demand.

This origin matters because contemporary use of the term still draws on it. When a current system is described as intelligent, the description does not only report that the system generates plausible text. It transfers the older claim that the system performs something resembling thought, reasoning, or understanding. The term retains a history of expectation that exceeds the current technical achievement.

Turing’s 1950 essay introduced a different framing. Turing opened with the question of whether machines can think, then set the question aside in favor of the Imitation Game. The test replaced the question of internal cognition with the question of external indistinguishability. If a machine could participate in a text-based exchange well enough that an interrogator could not reliably distinguish it from a human participant, the behavioral question would be answered.³

Turing’s substitution was reasoned. He treated the term “think” as philosophically unstable and proposed a disciplined proxy. The consequence was a shift in what counted as evidence for cognition. Observable success in linguistic performance came to function as a sufficient indicator. Once intelligence is framed around indistinguishable output, fluent surface behavior is granted evidential weight that the methodological substitution did not by itself justify.

Two lines emerge from the early history of the field. The Dartmouth proposal tied the term to the mechanization of intelligence. Turing tied evaluation to behavioral performance. The two lines do not coincide. The first concerns whether cognition can be structurally described and simulated. The second concerns whether a machine can perform in ways that defeat human discrimination. Contemporary large language models satisfy the second standard more clearly than the first.

The first major period of AI research remained closer to the Dartmouth ambition. Symbolic AI attempted to model reasoning through explicit rules, formal logic, and structured representations. Researchers encoded knowledge in systems that manipulated symbols according to defined procedures. Expert systems narrowed this approach into domain-specific rule structures. These systems performed within bounded tasks. They were also brittle. They struggled with context, ordinary language, common sense, and the open-ended complexity of human situations.⁴

The failure of symbolic systems is informative. The systems did not fail because the underlying ambition was incoherent. They failed because the world contains exceptions, tacit background, embodied habits, shifting purposes, and forms of relevance that cannot be listed in advance. A system could process symbols correctly and still fail to operate in the world in which symbols function. This established a working distinction. Symbol manipulation is a necessary condition for many kinds of computation. It is not by itself sufficient for the broader cognitive capacity the original term referenced.

The AI winters followed from this gap between projected and delivered capability. Funding contracted when systems failed to meet the expectations attached to the term. The term itself did not disappear. It remained available as the label for whatever future technology might restore confidence in the original program.

The technical restoration that followed did not come from a completed theory of machine understanding. It came from the statistical turn. From the 1990s into the 2000s, the field shifted from hand-coded rules to machine learning systems trained to detect patterns in data. The deep learning phase intensified the shift. Knowledge ceased to be represented primarily as explicit symbolic structure. It was distributed across high-dimensional mathematical spaces, optimized through training on large datasets. By the time Transformer-based language models became central, the technical paradigm had changed.⁵

The criterion of success changed with the paradigm. In symbolic AI, failure typically appeared as breakdown in a rule system. In statistical language modeling, success appears as plausible continuation. The criterion is performative rather than architectural. A system is rewarded when its output fits the surrounding linguistic context. The achievement is not trivial. It consists in producing form that human readers can take up as meaningful.

The term, however, did not change with the paradigm. A phrase introduced to name a program for the mechanization of intelligence remained in use as the name for systems whose primary competence is statistical pattern learning at scale. The new systems were more publicly impressive than most earlier symbolic systems. They could write, translate, summarize, classify, answer, imitate styles, and maintain extended coherent output. Their performance gave the older term renewed plausibility in public use.

That plausibility is misleading at the level of technical description. Large language models are not weak at producing language. They are highly capable producers of language. The technical achievement is statistical and generative. The system learns patterns in linguistic material and generates outputs that are coherent and context-sensitive. This achievement should be described in terms of its own structure. It is fluency, not the Dartmouth program of machine cognition.

The distinction matters for two reasons. The first is technical. Treating fluency as evidence of cognition imports unsupported assumptions about understanding, judgment, and reference. The second is interactional. Human readers are formed by ordinary language to treat coherent speech as the expression of intention, thought, and responsibility. When a model generates fluent language, the reader supplies the missing context. Fluency becomes the interface through which statistical generation is received as understanding. The reception is consistent with how humans process language. It is not consistent with what the system performs.

For these reasons the term “artificial intelligence” is not a neutral label. It maintains the conditions under which fluent output is treated as cognitive achievement. The older research aspiration is transferred onto a system whose operation does not satisfy it. The system receives a cognitive title before the technical description has earned it.

A more accurate description begins from what the system does. The system produces generated linguistic form, trained on the statistical structure of public language. This paper proposes the term artificial fluency for that operation. The term does not deny usefulness. It does not deny complexity or generative power. It denies that syntactic and statistical competence should be treated as cognition. Naming the system artificial fluency states the technical claim at the level at which the claim can be tested.

The historical arc can now be summarized. The term “artificial intelligence” was introduced as a conjectural name for a research program directed at the mechanization of intelligence. The symbolic era pursued that program through explicit representations and rules and encountered persistent limits. The statistical era changed the technical object. The inherited term remained. The LLM era produced systems whose public effect makes the older term appear plausible, although the technical achievement belongs to a different category.

The detachment between term and object is not only a matter of vocabulary. A descriptive term directs attention toward particular features of the object it names. It can encourage users to ask whether a system is intelligent before asking what kind of operation the system performs. It can encourage institutions to treat output as judgment, synthesis as understanding, and confident prose as authority. The following sections examine the technical, philosophical, political, and empirical consequences of this misalignment. The historical point is preliminary to those examinations. The term used for the present systems names a research program their operation does not satisfy.

2. What LLMs Actually Do

The historical point now becomes a technical one. If the term “artificial intelligence” has detached from the systems it names, the next step is to describe those systems without the assumptions the term imports.

Large language models are not autocomplete systems in any trivial sense. That description understates the scale and flexibility of their output. They are systems for generating statistically plausible linguistic continuations across many genres, tasks, and contexts. The capacity is real. The question is where the capacity is located. The answer is in fluency, understood as the production of coherent linguistic form from learned statistical relations.

A contemporary LLM is typically built on the Transformer architecture. It is trained on large corpora of text through self-supervised learning. The training objective is next-token prediction. A token is a machine-operable unit, usually a word fragment or subword unit. Given a sequence of prior tokens, the model estimates a probability distribution over the next token and is trained to reduce prediction error, generally through cross-entropy loss. During generation, the model samples or selects a token, appends it to the context, and repeats the process. This procedure is autoregressive: each output becomes part of the input for the next step.⁶

The Transformer architecture made the procedure more effective by allowing the model to weigh relations among tokens across a context. Attention is a mathematical procedure for assigning weight to elements in a sequence. A long context window preserves more prior text and supports coherence over longer exchanges. This should not be described as memory in the human sense, or as a horizon of understanding. It is a defined range of operable input.

Instruction tuning and reinforcement learning from human feedback adjust this process for human use. They increase the probability that the model produces outputs users find helpful, harmless, and conversationally appropriate. They reduce certain unwanted forms of output and produce a more cooperative interface. They can also produce output that sounds more deliberate than the underlying operation warrants, because responses have been trained into a socially recognizable shape. They do not change the category of operation. The model continues to generate by moving through token probabilities under a context window.

This technical description sits inside a prior transformation of language. Before the model operates, language has already been converted into writing, archive, corpus, token, vector, latent proximity, and generated output. The model does not begin from the world in which words first carried meaning. It begins from language after language has been written, stored, extracted, segmented, and made computable. Tokenization divides written language into units the system can process. Vectorization places those units into high-dimensional relations that preserve patterns of usage as calculable proximity.⁷

This explains the system's effective range. Human worlds have left traces in written language. Institutions, practices, conflicts, genres, disciplines, laws, and forms of life have deposited themselves into text. A model trained on that text learns relations that are genuinely informative. It produces outputs that orient a reader, expose a pattern, draft a sentence, translate a phrase, or assemble a plausible explanation. The output becomes meaningful for human readers because human readers return it to a world.

The same point clarifies the limits of the system. Distributional relations are not random. They are traces of usage, and usage originates in lives, institutions, practices, and histories. The system inherits the traces. It does not inherit the practical, bodily, historical, and normative conditions under which the traces were produced. A vector relation encodes a pattern left by meaning. The encoding is not the relation in which meaning arose.

This is the point at which the distinction between syntax and semantics becomes operative.

Syntax is the formal ordering of signs or tokens. Semantics concerns meaning, reference, intention, and relation to what the signs are about. Mid-century computationalism proposed that a sufficiently complex syntactic engine could function as a semantic engine. John Haugeland summarized the proposal: take care of the syntax and semantics may take care of itself. Searle’s Chinese Room argument denied the inference. A system can manipulate signs according to rules and generate responses indistinguishable from understanding without understanding what the signs mean. Block’s “Blockhead” pressed the same problem against behaviorism: identical outward performance does not establish identical internal process. Harnad’s symbol grounding problem made the issue exact. A closed symbol system requires an account of how symbols become meaningful for the system, rather than only for an interpreter outside it.⁸

These arguments do not establish that no future machine could understand. They establish that fluent output does not by itself establish understanding. The distinction is operative for LLMs because LLMs produce precisely the kind of output that invites users to infer understanding. The prose is continuous, responsive, and stylistically appropriate. The system answers in the social form of an interlocutor, while its operation remains statistical generation.

Bender and Koller state the boundary in terms specific to language models. They distinguish form from meaning. Form is observable linguistic material: text, sound, tokens, or other transmissible signs. Meaning requires a relation between linguistic form, communicative intent, and the world. Their Octopus Test illustrates the distinction. An octopus that intercepts signals between two humans may learn the statistical pattern of the exchange well enough to imitate one party. The octopus lacks access to the world the messages concern. When the situation requires grounded understanding of a new event, imitation fails.⁹

The LLM is not the octopus. The test isolates a structural condition. Training on form provides access to patterns in language. It does not by itself provide access to the world-involving relation that makes language meaningful. Scale increases the volume of form and the precision of the correlations. It does not change the source of the training signal.

The distinction between formal and functional linguistic competence specifies the same point in cognitive terms. Formal linguistic competence is the capacity to produce correct linguistic form: grammar, syntax, morphology, style, genre, and probable continuation. Functional linguistic competence is the capacity to use language in world-involved activity: reasoning, tracking situations, coordinating with others, acting on goals, and applying common sense. LLMs show high formal linguistic competence. Their functional competence is uneven because it depends on capacities that are not identical with linguistic form.¹⁰

This distinction protects the argument from a misreading. The claim is not that LLMs are weak at language. They are typically strong at language. The claim is that mastery of linguistic form is not identical with comprehension, judgment, or responsibility. Fluency is a real competence. It is also a bounded competence.

The boundary appears under specific empirical conditions. McCoy and colleagues argue that LLMs should be assessed in terms of the task they were trained to perform: predicting the next token over text data. Their shift-cipher experiments illustrate the point. When the correct decoded answer is a high-probability word sequence, performance improves. When the correct answer is statistically rare, performance declines, even though the logical operation is identical. A system executing a rule as a rule would not vary with the probability of the resulting sequence. A prediction system varies because probability is part of its path to output.

Planning exposes a related condition. A language model generates plan-shaped text. The generation does not establish that the model has constructed, tested, and revised a plan under constraints. Kambhampati distinguishes the appearance of planning from the execution of planning. Chollet adds a further criterion: intelligence as skill-acquisition efficiency. A model may perform well on many tasks because of extensive prior exposure. This differs from acquiring a new skill from limited experience and transferring it under unfamiliar conditions. Marcus and Mitchell address adjacent points: brittleness, weak systematic generalization, shallow analogy, gaps in common sense, and persistent limits at the level of meaning. These critics differ in framing. They converge on the point that fluent performance can mask fragility in the same places where users infer reasoning.¹¹

The evaluation metrics used for language systems reflect the same boundary. Perplexity measures how confidently a model predicts a sequence. BLEU and ROUGE compare overlap between generated text and reference text. BERTScore evaluates distributional similarity in contextual embedding space. These measures track confidence, overlap, or similarity. They do not establish truth, grounding, or answerability.¹²

The same boundary clarifies the phenomenon described as hallucination. A hallucination is not best described as a temporary loss of contact with reality. The description imports the wrong model. The system has no stable truth relation that could be lost. It generates fluent continuation under uncertainty. When the most plausible continuation is false, or when the prompt exceeds reliable evidence, the system may continue in the same confident form. Retrieval, tool use, verification layers, and human review reduce many failures. They do not close the gap between fluent continuation and grounded truth.¹³

Hallucination is therefore not the opposite of fluency. It is one outcome of the same operation. The mechanism that produces smooth continuation across open-ended contexts also produces continuation when the system should stop, qualify, or refuse. Human readers often experience fluent confidence as knowledge because ordinary language usually originates from speakers who can be asked for reasons, challenged, or corrected. The system produces the grammar of an answer without the conditions of answerability.

Several research programs propose that this boundary may be moved by future architectures. Emergence-oriented accounts argue that large models may develop internal representations rich enough to support new capacities. The literature on grokking shows that networks can move beyond memorization and learn general structures after extended training. Bengio’s System 2 program directs deep learning toward more deliberate, compositional reasoning. LeCun’s world-model program is more direct in its rejection of autoregressive LLMs and proposes predictive models of the physical world rather than text-only training. These positions do not weaken the present claim. Several confirm its boundary. LeCun’s proposal begins from the insufficiency of text prediction for grounded intelligence. Bengio’s program begins from the limits of fast pattern matching. The emergence claim more favorable to current models still requires a further argument: that high-dimensional internal representation has become semantic comprehension rather than increasingly precise modeling of linguistic and task structure.¹⁴

The most direct counter to the artificial fluency description is the industry shift toward agentic systems. Chained operations, verification loops, planning modules, and tool-mediated execution are presented as the bridge from fluent output to functional intelligence. The argument is that agentic scaffolding establishes what single-turn generation could not, because the system acts upon the world.

The reply is structural. Scaffolding does not change the kind of operation at any step. Each link in an agentic chain is a generation event. The plan is generated. The verification step is generated. The decision to call a tool is generated. The interpretation of the tool’s output is generated. The revised plan that follows is generated. The chain is longer than a single-turn exchange. The operation at each link is the same operation as the one that began the chain.¹⁵

A concrete example clarifies the structure. Consider a system trained such that the token “CANNON” is wired to a physical cannon, so that whenever the system generates that token, a shot is fired. The cannon fires. The shot occurs in the world. The system did not decide to fire. It produced a token. A wire converted the token into a physical effect.

The example generalizes. When an agentic system generates a token that triggers a database call, the database returns data. When it generates a token that initiates an API request, the request is sent. When it generates a token that approves a transaction, the transaction is processed. In each case, something occurs in the world. In no case does the system perform an additional category of operation. The category at each link is token generation. The wiring converts generated tokens into effects.

The distinction matters for what agentic systems can be said to do. The system can be described in two senses. In a thin functional sense, the chain checks values against sources, executes calls, and conditions subsequent tokens on returned data. In a thick sense, no link verifies, decides, or acts in the way a person verifies, decides, or acts. No link can be asked for reasons. No link bears the consequences of the effect produced. The grammar of deliberation is present in the generated tokens: “let me check,” “first I will, then I will,” “this requires further reasoning.” These are generated continuations. They are not the operations they describe.

This bears directly on responsibility. When the cannon fires and harm follows, the question of responsibility has a determinate answer. The system is not responsible. The token is not responsible. Responsibility belongs to those who built the system, wired the token to the cannon, deployed the configuration, and authorized its operation. If the same system is described as an autonomous agent that decided to fire, the responsibility appears to migrate toward the agent and away from those persons. If the system is described as a generator wired to an effect, responsibility remains where it originates.

The same structure applies across agentic deployments. A system that screens candidates is not making hiring decisions; it is generating tokens that condition downstream selections that humans authorized. A system that drafts a legal filing is not practicing law; it is generating tokens that an attorney is required to verify and sign. A system that suggests a clinical course is not treating a patient; it is generating tokens that a clinician must own as a recommendation before they enter care. In each case, the effect in the world is real. The category of operation that produced the effect is token generation. The chain of responsibility does not pass through the system.

Agentic scaffolding therefore extends reach without changing category. A system that calls APIs, queries databases, writes files, and triggers downstream processes operates on a wider surface of the world. It does so by producing fluent linguistic outputs at each step. The outputs are not less generation-shaped for being chained. They are not less ungrounded for being framed as actions. Describing the chained behavior as agency repeats the original misnaming at higher scale. The system does not fail to produce effects. The effects are produced through a sequence of generations, not through the work of a subject who can be asked for reasons, bear consequences, or undergo correction.

Future architectures may change the field. Systems with persistent memory, embodied feedback, causal world models, tool-mediated verification, planning modules, and stronger self-correction would require their own analysis. The present claim is narrower. Scale, longer context, retrieval, multimodality, and agentic scaffolding produce richer outputs and more convincing interactions. They do not by themselves establish experience, judgment, or responsibility.

The technical description of the present achievement is artificial fluency. LLMs transform written public language into tokenized and vectorized relations, then generate fluent linguistic form from those relations. The result is useful, powerful, and institutionally significant. It supports search, drafting, translation, comparison, coding, summarization, and analysis. It does not establish that the system understands its output, inhabits the world its words concern, or judges the consequences of what it produces.

The competence should be described at the level where it can be examined. LLMs achieve fluency: statistically trained, context-sensitive, generative command of linguistic form. Semantic comprehension, worldhood, responsibility, and judgment lie outside what that achievement establishes.

3. Fluency Without Experience

Section 2 described the technical operation of large language models. They generate fluent linguistic form from learned statistical relations. The next question concerns the human encounter with that form. The system produces tokens. The user does not encounter tokens. The user encounters an answer, a suggestion, a synthesis, a draft, a warning, or a piece of advice. Artificial fluency becomes consequential at this point, where generated form enters a situation that already matters to someone.¹⁶

The technical account is therefore incomplete without an account of reception. LLMs can lack semantic comprehension and still produce language that is received as understanding. The reception does not occur inside the model. It occurs in the relation between the fluent output and the user. Human beings are formed by ordinary language to expect that coherent speech originates with a speaker, that answers express an intention, and that confident prose belongs to someone who can be asked for reasons. Artificial fluency takes up this social grammar of language.

This is not a simple error by inattentive users. In ordinary life, the connection between fluency and understanding is usually sound enough to guide action. A colleague who gives a clear answer can be asked to explain it. A doctor who offers advice stands within a practice of licensing, evidence, and liability. A teacher who corrects a student can be challenged by the student or by the discipline. Speech normally belongs to relations of answerability. Artificial fluency interrupts that connection. It produces the linguistic form of answerability without the speakerly conditions that would make answerability possible.

The form of the chatbot interface reinforces the interruption. The exchange contains turn-taking, politeness, repair, and apparent patience. The system retains earlier inputs and returns them in later outputs. These features produce social continuity at the level of text. Continuity of text is not continuity of a life. Contextual response is not responsibility. The user encounters a form that resembles dialogue. The system remains a technical object that generates under constraints.

The result is an asymmetry. The user can be affected by the exchange. The system cannot be affected in kind.

The distinction between *Erlebnis* and *Erfahrung* clarifies the structure of the asymmetry. English uses one word, experience, for two different relations. *Erlebnis* denotes the lived event: the occurrence, the immediacy, the moment in which something happens to a subject. *Erfahrung* denotes formative undergoing: experience tested by time, negation, correction, memory, and revised understanding. *Erlebnis* names the event. *Erfahrung* names the subject changed by the event.¹⁷

A generated answer can constitute an *Erlebnis* for the user. Language arrives. A difficulty acquires form. An uncertain question becomes answer-shaped. The user is oriented, informed, or relieved. This event has practical value. It can begin inquiry, clarify a term, prepare a meeting, organize scattered material, or help a writer past an initial difficulty. The event does not become *Erfahrung* by occurring. It becomes *Erfahrung* only if the user returns through the output to the sources, evidence, people, practices, and world the output concerns.

The system itself produces no experience of either kind. It has no *Erlebnis*, since nothing is lived as occurrence. It has no *Erfahrung*, since it undergoes no formative correction. It does not suffer the failure of a first interpretation, become answerable for a claim, acquire tact through cases, retain an error as caution, or encounter another person as someone who can resist it. The system retains context, calls tools, retrieves documents, and generates revised outputs. These operations improve output. They do not constitute a subject who has undergone something.¹⁸

Generation is one operation. Undergoing is another. The distinction does not diminish the usefulness of generated language. It refuses the transfer of the user's reception into the system that produced the output. The user may have an experience in the course of using the model. The model does not have an experience in the course of being used.

The technical sequence in Section 2 supports the distinction. Generative AI begins after human beings have spoken, written, archived, argued, classified, and produced linguistic traces. Its order is writing, archive, corpus, token, vector, latent proximity, generated output. The sequence is consequential because the system begins from language after meaning has already become available as processable material. It does not begin from significance, care, embodiment, historical belonging, or responsibility. It begins from corpus.¹⁹

The output is therefore de-historicized without being empty. The historical relation among writer, reader, source, institution, conflict, and world has been converted into formal relation. A legal argument is converted into corpus. A poem is converted into token sequence. A professional report is converted into pattern. A person's career is converted into profile. Something is preserved in the conversion, since otherwise the output would be useless. Something is lost, since otherwise the output would constitute understanding.

The user supplies the missing context. This is user vivification. The model generates linguistic continuation. The user receives that continuation inside a task, fear, hope, deadline, ambition, obligation, or question. The model has no concern. The user is concerned. The model has no project. The user has one. The model has no responsibility for the answer. The answer enters a situation in which responsibility exists.²⁰

Vivification is more precise than anthropomorphism. Anthropomorphism denotes the attribution of human traits to a non-human entity. Vivification denotes the process by which generated language becomes experientially active because a user inserts it into a context in which the answer matters. A user can know that no person stands behind the output and still receive the output as if understanding had occurred. The interface does not need to deceive the user. It only needs to provide fluent language in the grammar of response.

The interface conditions help explain why fluent output is received with reduced critical resistance. Not-knowing is not only an information deficit. It is exposure: the user has not yet read, practiced, or thought enough to say what the matter requires. Artificial fluency reduces that exposure by providing language first. Where there is hesitation, it provides form. Where there is difficulty, it provides structure. Where there is silence, it provides continuation. This effect is not a private defect in the user. It belongs to the situation into which the technology arrives, a situation already shaped by speed, institutional pressure, and the demand for visible competence.²¹

The risk is therefore not limited to false output. False output can often be checked, challenged, or rejected. A more consequential risk is unowned truth. A generated answer can be accurate and still remain unappropriated. It can enter an email, a report, a classroom, a board deck, a legal memo, a clinical note, a hiring assessment, or a decision record without anyone having passed through the work by which the claim becomes answerable as their own.²²

Unowned truth is consequential because it does not present as failure. It presents as competence. The sentence is clean. The summary is accurate. The structure is useful. The problem becomes visible only when someone is asked to defend the claim, explain its route, name its conditions, account for its omissions, or revise it under pressure. If no one can do this, the truth has circulated without becoming knowledge in the responsible sense.

Verification is necessary but not sufficient. A person can confirm that a statement is correct and still not own it. Ownership requires appropriation: the return of generated language to source, history, practice, question, and responsibility. In scholarly work this means reading the cited text. In professional work it means returning the profile to the person, the brief to the organization, the risk to the evidence, or the recommendation to the one who must defend it. Generated output becomes material for *Erfahrung* only when it passes through such a return.²³

This return alters the status of the user's initial *Erlebnis*. The first arrival of language is immediate and relieving. The later return is slower and frequently less pleasant. It can show that the output was wrong. It can also show that the output was too smooth, too general, too detached from a conflict, too quick to reconcile a contradiction, or too confident about a matter that required hesitation. Where this occurs, the output has served inquiry because it has been resisted.

The institutional form of the same problem is more consequential. In organizations, generated language does not remain in the private experience of one user. It moves through workflows. It appears as summaries, notes, charts, recommendations, candidate assessments, meeting records, policy drafts, compliance narratives, and decision rationales. Each transfer reduces the appearance of provisionality. The generated origin recedes. The institutional authority increases.

This is institutional vivification. An organization treats generated language as if it had passed through the professional process whose form it imitates. The report has the form of analysis. The assessment has the form of judgment. The summary has the form of reading. The recommendation has the form of deliberation. Because each document can be copied, revised, and circulated, the initial *Erlebnis* of fluent arrival multiplies across the institution without the return through which *Erfahrung* would occur.²⁴

The problem is not that institutions use generated language. The problem is that institutional systems reward the visible product more readily than the formative passage behind it. A clean report travels faster than the hesitation that would complicate it. A confident summary travels faster than the unresolved question. A recommendation travels faster than the encounter that should have tested it. Artificial fluency fits this administrative tempo because it produces the outer forms institutions already recognize as work.

This is the connection between the technical claim and the political-economic claim. Fluency becomes commercially and institutionally valuable because it generates the outer forms of understanding at scale. The misnaming of this capacity as intelligence then produces a second effect. The system is not only used as a source of fluent material. It is received as a cognitive actor whose output warrants deference. The user's reception of apparent comprehension is converted into an institutional relation of trust.

The same structure applies to the responsibility analysis in Section 2. When generated output enters a workflow and an effect follows, the question of responsibility has a determinate answer. The system did not deliberate. The output did not decide. Responsibility belongs to those who deployed the system, designed the workflow, authorized the output, and gave it institutional force. Institutional vivification obscures this answer. It treats the document as if the work had been done. The document then circulates as if a responsible subject stood behind it.

4. The Political Economy of the Misnaming

A descriptive term organizes money, trust, purchase, liability, and policy. The misnaming of artificial fluency as artificial intelligence therefore has effects beyond vocabulary. It influences what receives investment, what enterprises buy, what users defer to, and where responsibility is located when generated output produces harm.

Section 3 described how fluent output is vivified by users and institutions. A generated sentence arrives as an answer. A generated summary enters a workflow. A generated assessment becomes part of a decision record. The term “intelligence” stabilizes this movement. It converts a linguistic instrument into something that can be sold, funded, and deferred to as a cognitive actor.

The term expands the imagined object. Described as artificial fluency, the technology has a large but bounded market. It includes drafting, summarization, translation, search, coding assistance, document comparison, workflow support, and other tasks in which generated language is useful. Described as artificial intelligence, the market becomes all cognitive labor. The software is no longer priced as a tool that assists workers. It is priced as an infrastructure layer that may replace workers, departments, professional processes, and decision systems.

This is the first political-economic effect of the term. “Intelligence” expands the total addressable market. It permits vendors and investors to move from software budgets to payroll budgets, from task assistance to labor substitution, from text generation to general cognition. If the technology were described as artificial fluency, its addressable market would remain within bounded software categories. Described as intelligence, it can be presented as a substitute for white-collar labor at large.²⁵

This helps explain the scale of post-2022 capital flows. Venture investment concentrated sharply in AI startups, and foundation model valuations detached from ordinary software multiples. Major foundation model providers were priced at levels not supported by current revenue. The valuations were supported by the projection that these companies would become basic utilities of a future cognitive economy.²⁶

The numbers are time-sensitive. The underlying mechanism is more stable than any particular figure. Capital is not flowing only toward improved document tools. It is flowing toward the projected replacement or reorganization of human cognitive work. The term “intelligence” makes that projection financially legible.

The same logic appears in the infrastructure debate. Sequoia’s “\$200B Question” and later “\$600B Question” frame the gap between the capital spent on chips, data centers, energy, and cloud infrastructure and the revenue that must follow to justify the expenditure. Goldman’s skeptical analysis makes a related point: large capital expenditure requires the technology to solve difficult problems reliably, not only to generate useful language at the edge of existing workflows. Acemoglu’s more modest productivity forecast gives the same debate a macroeconomic form. Each line of analysis identifies a gap between present utility and projected valuation.²⁷

The gap is bridged by the term. Artificial fluency can be useful without supporting a trillion-dollar infrastructure story. Artificial intelligence can support that story because it keeps the future open at the level of general cognition. The projected value is always one step beyond the current use case. Present systems draft and summarize. Future systems may reason, plan, manage, advise, decide, discover, and replace. “Intelligence” allows present fluency to borrow against future cognition.

The same effect appears in ordinary product language. AI systems are described as understanding, reasoning, thinking, knowing, hallucinating, and acting as agents. These verbs are not external to the product. They condition the user’s relation to it. A user told that a system “understands” a document relates to the output differently from a user told that the system has generated a statistically plausible transformation of the text.

“Hallucination” illustrates the pattern. In clinical use, hallucination denotes a subject with perception. Applied to a language model, the term implies that the system normally has a relation to reality and has briefly misperceived it. The accurate description is more limited: the system has generated plausible but ungrounded language. Each anthropomorphic term imports a cognitive state into a system whose operation is statistical generation and automated workflow execution.²⁸

The “stochastic parrots” critique and the older “wishful mnemonics” critique identify the same effect from inside computer science. A system named with cognitive verbs becomes harder to perceive as a system. The label functions as a claim about the object. Where a loop is named “understand,” the burden of proof is softened. Where a workflow is named an “agent,” its execution is described as action. Such terms may begin as shorthand. In commercial settings, shorthand becomes product ontology.²⁹

The pattern predates current LLMs. Early computers were promoted as electronic brains. Expert systems were sold as if codified rules could reproduce professional expertise. IBM’s Watson was marketed under the language of cognitive computing, particularly in medicine. In each case, cognitive naming sold a technology as closer to human judgment than its structure supported. Watson Health is a relevant precedent because it showed how a system could be framed as clinical cognition while struggling with the conditions of actual medical practice.³⁰

The current cycle differs in scale and interface. LLMs are more publicly persuasive because they produce fluent language in direct response to user input. Earlier systems often failed visibly. Contemporary systems fail inside a smooth answer. The surface is coherent even where the underlying relation to truth, context, or action is weak. This intensifies the political economy of naming. A brittle expert system could appear brittle. A language model can sound competent while it is wrong, incomplete, or outside the task boundary.

Enterprise buyers participate in this misnaming because they face their own incentives. Boards and executives are under pressure to demonstrate AI adoption. CIOs, CTOs, and other senior leaders are rewarded for appearing technically current. Consulting firms support this pressure with adoption narratives, investment forecasts, and organizational programs centered on AI readiness. Vendors then provide demonstrations under ideal conditions: a clean prompt, a fluent answer, a striking comparison, a polished result. The demonstration converts fluency into apparent reliability.³¹

The procurement problem begins when the demonstration becomes the procurement story. A company can buy intelligence and deploy fluency. It can announce transformation and use the tool for document generation. It can place the system into decision workflows because the institutional mandate requires AI adoption, rather than because the use case has been carefully bounded. When the system is introduced into messy internal environments, workers inherit the hidden burden of verification. The output is produced quickly. The verification becomes slow, less visible, and less rewarded.

This is the second political-economic effect. The term shifts the burden of proof. A text-generation tool must establish that its output is useful for a defined task. An intelligent system is granted broader deference. The user asks less often what the system has generated and assumes more often that the system has analyzed the matter. This is consequential in institutions because generated language circulates with declining visibility of origin. A draft becomes a note. A note becomes a file. A file becomes a record. A record becomes evidence that a process occurred.

The institutional vivification described in Section 3 explains the trajectory. Generated language enters documents, dashboards, reports, presentations, recommendations, and decision records. Each circulation increases authority and reduces visibility of origin. The organization functions as an *Erlebnis* multiplication engine: language arrives, is edited, forwarded, and authorized. The return to source, case, and encounter is not required by the workflow.³²

At this point the naming question becomes a liability question. The cannon example in Section 2 establishes the structural answer. When a system generates a token wired to an effect, the effect occurs in the world. The system did not decide. Responsibility belongs to those who built the system, wired the token to the effect, and authorized the configuration. The same structure applies in institutional settings. When generated output enters a workflow and produces harm, responsibility belongs to those who deployed the system, designed the workflow, and gave the output force.

The term “intelligence” obscures this structure. When a system is described as intelligent, harmful output can be presented as the action of a semi-autonomous agent. Responsibility appears to disperse across developer, vendor, deployer, user, model, data, and workflow. The system acquires the position of a juridical intermediary between the corporate deployer and the injured party.³³

The contrast is straightforward. Where the description is “an AI agent decided to fire,” the deployer stands behind the agent. Where the description is “a token generator wired to a cannon,” the deployer stands in front of the cannon. The renaming proposed here does the same work in hiring, credit, clinical advice, and legal filings. The wire produces the effect. The deployer strung the wire.³⁴

Regulatory frameworks are already addressing this problem. The EU AI Act classifies employment-related AI systems as high-risk. GDPR Article 22 concerns automated decision-making. The SCHUFA decision treated rubber-stamping of an algorithmic score as automated decision-making rather than meaningful human oversight. The problem is not the absence of human presence. The problem is the weakness of human ownership. A person can be “in the loop” and function as a legitimating surface for machine output.³⁵

The term artificial fluency would not resolve regulation by itself. It would clarify the object under regulation. A generated recommendation is not a decision. A generated assessment is not judgment. A generated reason is not responsibility. The professional standard follows from this. The question is not whether AI was used. The question is whether a responsible professional can defend the claim as their own judgment. Tool use is not the boundary. Answerability is the boundary.³⁶

The labor debate is distorted by the same naming problem. Where the system is described as intelligence, the debate concerns replacement of human cognition. Where the system is described as fluency, the debate concerns task allocation: which tasks are primarily linguistic production, transformation, or retrieval, and which tasks require situated judgment, responsibility, and encounter. These are different questions with different policy implications.

The empirical labor literature supports a moderate account. Generative systems often complement workers at the task level rather than replacing occupations wholesale. A customer-support study involving more than 5,000 agents found productivity gains, with larger gains for novice and lower-skilled workers than for highly skilled workers. The pattern is consistent with skill compression: baseline output improves, particularly in work containing repeatable communication patterns, while complex cases continue to require human oversight and judgment.³⁷

The distinction matters because artificial fluency alters the structure of skill formation. It raises the floor of competent-looking output. It enables novices to produce the language of a field sooner. It can also conceal whether the novice has formed the judgment required to evaluate the output. This is the risk of expertise without formation: where users no longer practice summarizing, arguing, reading, or writing, the capacities behind those acts may attenuate.³⁸

This does not require a general anti-technical position. Some fluency-class tasks can be accelerated safely. Some friction is waste. Some tools support apprenticeship by making examples easier to compare, drafts easier to critique, and sources easier to find. The error begins when fluency-class acceleration is treated as judgment-class substitution. Translation, summarization, code assistance, and document preparation are not the same kind of task as diagnosis, sentencing, hiring, strategy, or professional recommendation.

The geopolitical frame extends the same inflation to state level. Described as artificial fluency, the technology remains strategically significant. It supports propaganda, cyber operations, translation, intelligence summarization, software development, and institutional coordination. These are serious capacities. Described as artificial intelligence, the technology is positioned within an arms-race narrative about artificial general intelligence, superintelligence, and state survival. The narrative directs policy attention away from present harms and toward speculative future scenarios.³⁹

This does not render every safety concern imaginary. A system that can generate persuasive language at scale can produce political and social harm. A system integrated into cyber or military workflows is consequential. The issue is proportionality. The term “intelligence” encourages policy to imagine a rival mind emerging from the machine. The term “fluency” keeps attention on the actual capacities: generation, synthesis, imitation, classification, and automated action under human-designed systems.

The standard objection is that AI is shorthand. The term is familiar, convenient, and already embedded in law, markets, media, and public usage. Renaming does not alter model weights, training data, interface, or output. The objection is accurate about the mathematics and incomplete about the social relation. Terms shape budgets. They shape procurement thresholds. They shape what executives understand themselves to be buying. They shape what users believe they are facing. They shape legal arguments about autonomy and responsibility. They shape policy categories and public expectations.

“Artificial fluency” functions as an economic and political demotion. It lowers the claim from autonomous cognition to generated linguistic capacity. The demotion is the point. A tool remains useful after demotion. A language engine remains valuable when stripped of cognitive attribution. The objective is not to make the technology smaller than it is. The objective is to make the claim accurate enough that responsibility can return to the persons and institutions that design, sell, buy, deploy, and authorize it.⁴⁰

5. Empirical Harms in Judgment-Class Domains

The category mistake becomes visible in the empirical record. Harms concentrate where institutions deploy a system of fluency, prediction, classification, or pattern matching to perform situated judgment. The cases differ by domain, technology, and legal setting. They share a structure. A system produces an output that appears cleaner than the human process it enters. The output is faster, more consistent, and easier to circulate. It then receives institutional force before anyone has owned it as judgment.

A judgment-class domain is one in which the task cannot be reduced to ranking, prediction, pattern matching, or fluent synthesis without loss of what makes the task responsible. The domain requires interpretation of persons, contexts, histories, conflicting goods, and consequences. Fluency-class tasks align with generation, summarization, translation, matching, or bounded pattern recognition. Judgment-class tasks require situated reasoning, contextual interpretation, and accountability under contested goods.⁴¹

Four conditions identify the judgment-class problem. First, the target variable is unstable or contested. “Good candidate,” “safe defendant,” “appropriate treatment,” “creditworthy applicant,” and “student merit” are not simple natural kinds. Second, the data are historically loaded. Records of hiring, policing, grading, treatment, and credit do not only describe the past. They contain prior institutional choices. Third, the output affects a person who may not be able to inspect, contest, or correct the pathway by which they were classified. Fourth, the decision requires ownership by someone who can give reasons. Where these conditions hold, fluency and prediction can assist judgment. They cannot replace the process through which judgment becomes answerable.

This does not idealize human decision-making. Human judgment is often biased, noisy, self-protective, hurried, and socially narrow. The empirical question is whether automation corrects those defects or gives them a faster institutional form. In judgment-class domains, the record shows the second pattern. The system does not escape history. It learns from historical data, proxy targets, institutional records, prior decisions, and existing categories. It returns those inputs as outputs with the appearance of objective assessment.

Senior executive hiring is the clearest case because the boundary is visible inside the practice. The author writes from within this practice as the founder of an executive search firm, and the position should be stated. The argument applies to the firm’s own work. Where representations are returned to person, organization, timing, and reasons, the practice can be defended as situated judgment. Where they are not, the same critique applies regardless of who produced the output.

A profile shows title, employer, dates, responsibilities, sector, geography, public achievements, and visible career sequence. These traces are real. No search can begin without representation. The object of senior search is a person entering a possible organizational world. The practice begins with representations and becomes substantive when those representations are returned to the person, the organization, the timing, and the reasons. A trace can be processed. A person must be encountered. The encounter is what distinguishes search from matching.⁴²

For this reason senior hiring cannot be reduced to profile matching. The relevant question is whether this person makes sense for this organization, under these conditions, at this moment. At senior level, capability is typically a threshold rather than a differentiator. The decisive questions concern judgment, authority, identity, motivation, value, timing, and fit with a concrete organizational world. A model can support comparison, typification, retrieval, ranking, and summary. Singularization requires encounter, narrative judgment, ethical attention, temporal understanding, and responsibility.⁴³

Algorithmic hiring systems first fail by converting a contested judgment into a proxy target. A company rarely has full longitudinal data on the applicants it did not hire. It has data on current and past employees, performance reviews, retention, promotion, sales, credentials, and managerial evaluations. These data are shaped by prior hiring, promotion, exclusion, and survivorship. When a system learns from them, it does not identify the structure of merit. It learns the history of who was selected, retained, evaluated, and made legible as successful.

The Amazon recruiting tool remains the standard example. The system was trained on a decade of resumes submitted to Amazon and learned to penalize signals associated with women, including the term “women’s” in phrases such as “women’s chess club captain.” The tool was withdrawn after developers identified that it systematically downgraded female candidates. The structural point is general. Removing explicit demographic variables does not eliminate protected categories, since models can reconstruct them through proxies. Historical success data can preserve the exclusion the tool was intended to overcome.⁴⁴

Proprietary hiring vendors add a second problem: opacity. Video assessment, game-based assessment, and behavioral scoring systems claim to measure qualities such as grit, openness, decisiveness, communication, or fit. The validation procedures, training data, and weighting systems often remain inaccessible to candidates and difficult for employers to inspect. HireVue drew criticism for facial analysis and later retired that feature after public and regulatory pressure. Biometric and behavioral signals remain consequential where vocal intonation, speech patterns, reaction times, disability, accent, or neurodivergence can become evaluation inputs.⁴⁵

Employment law has begun to address this. EEOC guidance treats algorithmic resume screens, chatbot filters, and video evaluation tools as selection procedures subject to disparate impact analysis. The four-fifths rule provides one diagnostic measure: where the selection rate for a protected group falls below eighty percent of the selection rate for the majority group, the tool is flagged for adverse impact analysis. The employer remains responsible when a third-party vendor built the tool. This is a legal expression of the broader claim. An institution cannot disown a judgment-class output because the output was produced by software. The cannon example from Section 2 applies. The vendor built the wire. The employer connected it to a hiring decision. The candidate is the person harmed when the wire produces an effect. Responsibility runs to those who deployed and authorized the system.

The empirical record also shows that algorithmic assessment changes candidates before the system has assessed them. Goergen and colleagues identify an AI assessment effect. Across twelve studies with 13,342 participants, candidates who knew they were being evaluated by AI emphasized analytical traits and downplayed intuitive and emotional traits. The mechanism is direct: candidate nuance passes through awareness of AI evaluation, behavioral adaptation follows, and the output becomes a flattened analytical profile.⁴⁶

The effect is consequential in executive hiring. Senior leadership is not reducible to analytical performance. It involves judgment under pressure, relation to authority, trust, timing, moral courage, conflict, self-knowledge, and capacity to inhabit a particular organization. Where candidates adjust self-presentation toward what they understand machines to reward, the assessment does not observe the market. It shapes it. The system produces a narrower candidate world and then evaluates the narrower world as if it were the original one.

The same pattern appears in the pipeline literature. Studies of executive search show that gender disadvantage often occurs early in the funnel. Women may be less likely to be advanced to initial interviews, even where they face no statistical disadvantage at later client stages. Access to encounter is already uneven. Algorithmic screening trained on historically bent pipelines automates non-encounter. The candidate is not approached, not heard, and not given the opportunity to make reasons visible.⁴⁷

The portability of star performance deepens the point. Groysberg's work shows that star professionals frequently lose performance after moving organizations, particularly when they move alone or into firms with weaker supporting capabilities. Prior success was not only an individual trait. It was embedded in networks, resources, timing, colleagues, and institutional context. Profile matching abstracts the person from the world in which the success occurred. Senior search must ask how prior capability depended on a world and whether it can become real in another one.⁴⁸

Medicine extends the boundary into higher-stakes settings. Watson Health was marketed as cognitive computing for clinical judgment. Its flagship oncology project at MD Anderson ended after four years, \$62.1 million in fees and related costs, and no use in patient treatment. The system did not reconcile structured medical literature with messy electronic medical records. Later investigations identified unsafe and incorrect treatment recommendations in Watson for Oncology, frequently conflicting with local oncologists' situated judgments.⁴⁹

The Watson case does not establish that medical AI is useless. Bounded pattern recognition can be valuable in clinical settings where data are well defined and the task is narrow. The failure appears when synthesis of medical literature is treated as treatment judgment. Oncology does not concern a generic patient in a clean knowledge base. It concerns this patient, with this record, disease trajectory, comorbidity, history, preference, institution, family, and risk. A fluent recommendation can sound clinically authoritative and miss the case.

Patient-facing triage chatbots show the problem at the interface. Babylon Health promoted its triage system as matching or exceeding human general practitioners. Clinical criticism identified unsafe advice, including missed heart attack symptoms and excessive escalation for minor symptoms. The issue was not only error. The issue was corporate defense of a proprietary system at the point where patient safety required transparent correction.⁵⁰

Medical overtrust intensifies the risk. Studies have found that participants rate AI-generated medical replies as more trustworthy, valid, or complete than physician responses, even where low-accuracy and potentially harmful information was inserted. Fluency and authority of tone reduce skepticism. In clinical settings, artificial fluency can acquire the force of care without the relation, examination, liability, and professional formation that make care answerable.⁵¹

The clinical setting also clarifies how misnaming alters responsibility. Where a model is described as an intelligent medical assistant, a physician may be pressured to consult it, defer to it, or explain departure from it. Where it is described as artificial fluency, the output is generated material that must be clinically owned before it enters care. The difference is consequential because medicine is already accelerated and documentation-heavy. A fluent recommendation can save time at the surface while transferring the substantive burden to verification, correction, and professional risk. The cannon example applies. The model generates a recommendation. The recommendation is wired into a clinical workflow. The patient is the person who absorbs the effect. Responsibility belongs to the clinician, the institution, and the vendor who placed the system into care.

Law shows two distinct harms. The first is fluent fabrication. In *Mata v. Avianca*, attorneys used ChatGPT to draft a legal filing. The model produced judicial opinions, quotations, citations, and even airline names that did not exist. The attorneys submitted the material without adequate verification and were sanctioned. The case shows what happens when a generative model is treated as a legal research system rather than as a source of generated text requiring source-level confirmation.⁵²

The second legal harm is statistical injustice. COMPAS risk scores have been used in criminal justice settings to estimate recidivism risk. ProPublica's analysis identified racial asymmetry in error rates. Black defendants were more likely than white defendants to be classified as high risk and not reoffend. The deeper issue is the impossibility result in algorithmic fairness. Where base rates differ across groups, a model cannot satisfy all common fairness criteria at once. The selection of which fairness condition to optimize is a political and legal decision, not a mathematical one.⁵³

These legal examples are different in form. One concerns generated legal text. The other concerns predictive classification. They belong together because both replace answerable interpretation with an output that appears more settled than the human task permits. Law requires reasons that can be challenged, sources that can be inspected, and institutional responsibility for the application of norms. A fabricated case breaks that structure openly. A risk score breaks it more quietly by concealing a normative trade-off inside a number.

Credit and consumer finance show how opacity weakens accountability. The Apple Card controversy began when customers reported substantially different credit limits for men and women in the same household. The New York Department of Financial Services did not find intentional discrimination or broad disparate impact in the population it reviewed. The supported critique is more limited: frontline representatives could not explain the decisions to consumers. The model could be explained to regulators after the fact. The customer-facing relation was governed by “the algorithm” as an opaque authority. The substantive harm in the case was the failure of explainability at the point of contact, rather than a finding of statistical discrimination.⁵⁴

Air Canada shows the related problem of corporate disavowal. In *Moffatt v. Air Canada*, a website chatbot gave false information about bereavement fares, telling a customer that a refund could be requested after booking. Air Canada later denied the refund and argued that the chatbot was a separate legal entity, or that the customer should have consulted the static terms elsewhere. The tribunal rejected that defense and held the company responsible for the automated output it placed in its commercial interface. The case states the accountability boundary directly. A corporation cannot deploy fluent language in customer service and then disown it when the language causes harm. The chatbot is the wire. Air Canada placed the wire in front of the customer.⁵⁵

Education shows both aggregate injustice and probabilistic suspicion. The 2020 UK A-level grading scandal occurred when Ofqual used a statistical standardization model after exams were cancelled. The model used school historical grade distribution, teacher rank within cohort, prior GCSE attainment, and a small-cohort exception. Nearly 40 percent of teacher-predicted grades were downgraded. Private schools benefited disproportionately from the small-cohort exception. Students in historically lower-performing state schools were constrained by institutional history rather than evaluated as individuals. More than 20,000 students lost university placements before the model was withdrawn.⁵⁶

The structure is clear. The algorithm substituted institutional past for individual assessment. A student was not evaluated as this student under these conditions. The student was evaluated in part as an instance of a school's historical pattern. The model did not standardize grades. It converted social history into an individual fate.

AI plagiarism detection produces the same problem through suspicion. Traditional plagiarism tools match text against known sources. AI detectors infer authorship through probabilistic markers such as low perplexity and low burstiness. Real-world false-positive problems have been documented in Turnitin's AI detector, including the company's later caution about scores under 20 percent and concerns about sentences near actual AI writing. Independent research also identified elevated risk for non-native English speakers, whose simpler or more predictable sentence structures could be flagged as machine-generated. Students then face a structural presumption of guilt based on a probability score they cannot meaningfully disprove without intrusive documentation of their writing process.⁵⁷

The cases also show that harm does not require full automation. A human can remain formally present while the output governs the practical decision. The assessor glances at a score. The physician accepts a suggested reply. The lawyer copies a citation. The representative repeats an opaque denial. The teacher treats a detector score as sufficient suspicion. In each case, human presence functions as a procedural sign of oversight while the path of judgment has already narrowed. The phrase "human in the loop" is therefore insufficient by itself. The relevant question is whether the human can understand, challenge, revise, and own the output.

The scale of deployment changes the form of the error. A single bad judgment can be corrected by appeal, dialogue, or second opinion, at least in principle. A model embedded in an institutional workflow can reproduce the same misclassification across thousands of cases before the pattern becomes visible. The harm then appears as procedure rather than as discretion. No one needs to intend exclusion for exclusion to be stabilized. No one needs to act with malice for a person to be denied encounter, care, credit, education, or legal standing by a system whose categories were never adequate to the case.

Consistency is not judgment. The attraction of automated systems is understandable. Human decisions are noisy. A tired assessor, an impatient doctor, an overworked judge, a defensive bank, or a biased teacher can produce harm. Automated systems offer consistency. The empirical record shows that consistency assists only when the target is sound and the task is bounded. Where the target is historically biased, contested, or context-dependent, consistency multiplies error. The output appears orderly because the same mistake is made the same way.

This is the noise and bias problem. Human judgment may contain random scatter. Algorithms reduce that scatter by enforcing the same rule across cases. Where the rule is built from bad proxies, damaged history, or an inadequate target variable, the reduction of noise becomes the multiplication of bias.

The theoretical literature gives this pattern a broader form. Pasquale's black-box critique concerns hidden metrics and due process. O'Neil's analysis concerns models that encode opinions in mathematics at scale. Eubanks's work concerns the use of technical systems to manage vulnerable populations through surveillance and triage. These critiques differ. They converge on algorithmic flattening: the conversion of a human situation into a computable matrix that removes the context required for just judgment.⁵⁸

The conclusion is bounded. Artificial fluency and related predictive systems work effectively where the task fits the system. Protein folding, radiology under defined conditions, industrial defect detection, translation with caveats, code completion, and document summarization are cases where the target is bounded, the output is the product, or a ground truth is available. These uses are not the problem. They are why the technology is consequential.⁵⁹

The problem begins when a system that can generate, classify, summarize, or predict is treated as an epistemic agent. In that moment, a draft becomes a recommendation, a score becomes a judgment, a pattern becomes a person, a statistical regularity becomes a reason, and a fluent sentence becomes institutional authority.

The cannon example states the result in general form. The system produces a token. The token is wired to an effect. Where the effect harms a person, the question of responsibility has a determinate answer. Responsibility belongs to those who built the system, designed the workflow, authorized the output, and placed it in front of the affected person. The misnaming of artificial fluency as artificial intelligence allows that answer to be obscured. The renaming proposed here restores the answer.

6. Adopting “Artificial Fluency”

The proposal follows from the preceding analysis. It is limited in form and consequential in effect. The paper has not argued that contemporary systems are useless, trivial, or incapable of changing work. It has argued that the inherited term assigns them the wrong status. “Artificial intelligence” grants cognitive standing to a technical operation whose central achievement is fluent generation.

“Artificial fluency” describes the same systems at the level where the achievement can be examined. They generate linguistic, code, and media output from learned statistical relations. They support drafting, summarization, translation, classification, comparison, search, coding, and assistance. These are real capacities. The error begins when these capacities are treated as semantic comprehension, practical wisdom, or judgment. The proposed term does not reduce the technology. It reduces the claim made on its behalf.⁶⁰

The term “fluency” is precise because it names form. In ordinary use, fluency denotes smooth, competent, context-sensitive expression. In computational settings, the same concept is operationalized through confidence, probability, overlap, and distributional similarity. A fluent system produces text that reads well, follows genre conventions, and responds plausibly to a prompt. This is why the systems are useful. It is also the source of the misnaming risk. The surface of competence supports the inference of understanding.⁶¹

“Artificial fluency” therefore functions as a descriptive constraint on that inference. It does not state that the output is false. It does not state that the output should be ignored. It states that generated language has not become knowledge, judgment, or responsibility because it has arrived in coherent form.

The constraint matters because the user does not encounter a probability distribution. The user encounters an answer. The answer can relieve confusion, organize a task, prepare a note, or provide language where there was hesitation. Section 3 described this as the desire to have already understood: the wish to arrive at the far side of not-knowing without undergoing the formation that would make understanding one’s own. Artificial fluency intensifies that desire by providing competence-shaped language before the user has necessarily passed through reading, inquiry, encounter, or judgment.⁶²

The descriptive constraint places a small delay between output and uptake. It directs the user to return the generated language to sources, evidence, practice, person, and world. This return is not editing. It is reappropriation. A generated claim may be useful material for thought. It must still be made answerable by someone who can explain why it is being used, what supports it, what it omits, and what follows from it.⁶³

The professional version of the same boundary is the *Erlebnis* and *Erfahrung* distinction from Section 3. AI produces no *Erfahrung*. It does not undergo cases, acquire tact, suffer correction, or become responsible for a claim. The user's encounter with output may be an *Erlebnis*, a lived event of arrival. It becomes *Erfahrung* only where the user moves beyond the output into evidence, encounter, correction, and practice. Professional judgment requires this passage because judgment is not the possession of a representation. It is an owned claim under conditions of responsibility.⁶⁴

The renaming is therefore not cosmetic. A descriptive term conditions the social relation to the system. "Artificial intelligence" presents the system as a cognitive actor. "Artificial fluency" presents it as a linguistic and pattern instrument. The difference affects procurement, user trust, legal responsibility, and policy. A fluent tool produces material that must be owned. An intelligent agent appears to own part of the action.

The political economy of the previous term is built on this ambiguity. "Intelligence" expands the market from software assistance to cognitive labor substitution. It supports anthropomorphic verbs: the model understands, reasons, knows, hallucinates, acts. It allows vendors to sell deference rather than utility. It permits institutions to displace accountability when outputs harm people. The output appears to originate with an intelligent intermediary rather than with a system designed, sold, deployed, and authorized by human actors.⁶⁵

“Artificial fluency” removes the intermediary. It returns the output to the chain of human responsibility. The cannon example from Section 2 states the principle in general form. The system generates a token. The token is wired to an effect. The effect occurs. Responsibility belongs to those who built the system, strung the wire, deployed the configuration, and authorized its operation. Renaming does not resolve liability by itself. It clarifies where liability begins. A generated assessment is not a judgment. A generated reason is not responsibility. A generated recommendation is not a decision until someone owns it as such.

The empirical record described in Section 5 supports this distinction. Artificial fluency operates effectively where the task is bounded and the output is itself the product: translation, summarization, drafting, coding assistance, document comparison, and pattern recognition with clear ground truth. These are fluency-class tasks. The system’s competence matches the task because the task requires transformation, generation, retrieval, or bounded classification.⁶⁶

Judgment-class tasks are different. They concern persons and consequences under conditions in which the target is contested, historical, contextual, and value-laden. Hiring, medicine, law, credit, education, public administration, and professional assessment require more than fluent output or statistical prediction. They require reasons that can be challenged and revised. They require someone who can answer for the decision. The harms described in the previous section appear where fluent or predictive systems are granted institutional authority in these domains.⁶⁷

Policy should begin from this classification. The issue is not whether a tool uses machine learning, generative modeling, retrieval, or automation. The issue is whether the system is being used for a fluency-class task or placed in a judgment-class role. A summarizer used by a lawyer to organize documents is one kind of use. A system that fabricates citations and is treated as legal research is another. A model that drafts patient instructions for clinician review is one kind of use. A system whose advice substitutes for clinical judgment is another. A tool that helps map candidates is one kind of use. A score that prevents encounter is another.

This distinction also revises the meaning of oversight. A human “in the loop” is an insufficient standard where the person only approves output they cannot understand, defend, or revise. The stronger standard is ownership. Ownership requires that a responsible person can state the grounds of the claim, identify uncertainty, explain rejected alternatives, and bear the consequences of use. Human presence is procedural. Human ownership is substantive.

This standard does not require refusal of the technology. It requires formation-preserving use. Professionals require new skills: prompting, constraining, verifying, comparing, auditing, critiquing, and reappropriating generated outputs. These are real skills. They do not replace domain formation. They depend on it. A professional who lacks the relevant formation may not identify what the output has omitted, smoothed, exaggerated, or invented.⁶⁸

The point applies sharply to novices. Existing experts can often use artificial fluency effectively because they know where the boundary lies. They identify when a summary has flattened the issue, when a draft has concealed uncertainty, or when a recommendation has failed the case. Novices may receive polished language before they possess the judgment required to resist it. A practice that automates all peripheral tasks may also remove the path by which newcomers learn salience, evidence, proportion, and responsibility.

AI literacy is therefore insufficient where it means only facility with tools. AI literacy without professional formation is technical fluency without responsibility. The relevant question for institutions is not how much output can be generated, but which uses preserve the practices through which people become capable of judging the output. Some friction is waste. Some friction is formation. Artificial fluency must not be allowed to remove the second under the pretext of removing the first.⁶⁹

The likely objection is that “AI” is shorthand. The term is familiar, convenient, and already embedded in law, markets, media, and public usage. Renaming does not change model weights, training data, or inference. The objection is accurate about the mathematics and incomplete about the social relation. Shorthand still shapes budgets, deference, procurement, liability, regulation, and trust. It conditions what users believe they are facing.

The paper therefore does not call for ending deployment. It calls for accurate placement. Artificial fluency should be used where fluency, retrieval, transformation, comparison, and bounded pattern recognition are the task. It should be constrained where output affects persons in judgment-class domains. It should be named in a way that prevents fluency from acquiring the authority of understanding.

Artificial fluency can produce language. It cannot make that language answerable. That work remains with the human beings and institutions that use it.

Notes

¹ John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” August 31, 1955, 2, accessed May 21, 2026, <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. The quotation in the text follows the proposal’s opening conjecture.

² McCarthy, Minsky, Rochester, and Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” 1 to 13. See the proposal sections on information theory, neural nets, abstraction, randomness and creativity, and formal languages.

³ Alan M. Turing, “Computing Machinery and Intelligence,” *Mind* 59, no. 236 (1950): 433 to 460. For a secondary philosophical account of the Turing Test and its later interpretations, see Graham Oppy and David Dowe, “The Turing Test,” *Stanford Encyclopedia of Philosophy*, accessed May 21, 2026, <https://plato.stanford.edu/entries/turing-test/>.

⁴ For the symbolic period, expert systems, and the limits of rule-based AI, see Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (New York: Pantheon, 2019), chaps. 1 to 3; “Beyond deep learning,” *Proceedings of the National Academy of Sciences*, accessed May 21, 2026, <https://www.pnas.org/doi/10.1073/pnas.2214148119>. For the historical label “AI winter,” see “AI Winter: Understanding the Cycles of AI Development,” DataCamp, accessed May 21, 2026, <https://www.datacamp.com/blog/ai-winter>.

⁵ Marcus and Davis, *Rebooting AI*, chaps. 2 to 4; Tom McCoy, Shunyu Yao, Dan Friedman, Mathew Hardy, and Thomas L. Griffiths, “Embers of Autoregression Show How Large Language Models Are Shaped by the Problem They Are Trained to Solve,” *Proceedings of the National Academy of Sciences* 121, no. 8 (2024): e2322420121; Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell, “On the

Dangers of Stochastic Parrots: Can Language Models Be Too Big?” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York: ACM, 2021), 610 to 623.

⁶ McCoy et al., “Embers of Autoregression,” e2322420121; Bender et al., “On the Dangers of Stochastic Parrots,” 610 to 623; Long Ouyang et al., “Training Language Models to Follow Instructions with Human Feedback,” in *Advances in Neural Information Processing Systems* 35 (2022), 27730 to 27744; Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei, “Deep Reinforcement Learning from Human Preferences,” in *Advances in Neural Information Processing Systems* 30 (2017).

⁷ Alessio Montaruli, *The Vectorized Afterlife of the They: AI, Heidegger, and the De-historicization of Meaning* (2026), Introduction and chaps. 4 to 5. The sequence writing, archive, corpus, token, vector, latent proximity, generated output is developed there as the technical-philosophical order of generative AI.

⁸ John Haugeland, “Semantic Engines: An Introduction to Mind Design,” in *Mind Design: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, MA: MIT Press, 1981); John R. Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences* 3, no. 3 (1980): 417 to 424; Ned Block, “Psychologism and Behaviorism,” *The Philosophical Review* 90, no. 1 (1981): 5 to 43; Stevan Harnad, “The Symbol Grounding Problem,” *Physica D* 42, nos. 1 to 3 (1990): 335 to 346.

⁹ Emily M. Bender and Alexander Koller, “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Stroudsburg, PA: Association for Computational Linguistics, 2020), 5185 to 5198.

¹⁰ Kyle Mahowald, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko, “Dissociating Language and Thought in Large Language Models: A Cognitive Perspective,” *Trends in Cognitive Sciences* 28 (2024). See also the preprint version, arXiv:2301.06627.

¹¹ McCoy et al., “Embers of Autoregression”; Subbarao Kambhampati, Karthik Valmeekam, Lin Guan, Mudit Verma, Kaitlyn Stechly, Siddhant Bhambri, Lucas Paul Saldana, and Sarath Sreedharan, “Can Large Language Models Reason and Plan?” arXiv:2403.04121 (2024); Subbarao Kambhampati et al., “LLMs Can’t Plan, But Can Help Planning in LLM-Modulo Frameworks,” arXiv:2402.01817 (2024); François Chollet, “On the Measure of Intelligence,” arXiv:1911.01547 (2019); Marcus and Davis, *Rebooting AI*; Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (New York: Farrar, Straus and Giroux, 2019).

¹² For perplexity as a language-model confidence measure, see McCoy et al., “Embers of Autoregression.” For overlap and similarity metrics, see Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, “BLEU: A Method for Automatic Evaluation of Machine Translation,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (2002), 311 to 318; Chin-Yew Lin, “ROUGE: A Package for Automatic Evaluation of Summaries,” in *Text Summarization Branches Out* (2004), 74 to 81; Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi, “BERTScore: Evaluating Text Generation with BERT,” arXiv:1904.09675 (2019).

¹³ Bender et al., “On the Dangers of Stochastic Parrots,” 610 to 623; Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli, “Hallucination Is Inevitable: An Innate Limitation of Large Language Models,” arXiv:2401.11817 (2024); Lei Huang et al., “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions,” *ACM Transactions on Information Systems* 43, no. 2 (2025): 1 to 55.

¹⁴ For emergence and grokking, see Jason Wei et al., “Emergent Abilities of Large Language Models,” *Transactions on Machine Learning Research* (2022); Alethea Power, Yuri Burda, Harri Edwards, Igor Babuschkin, and Vedant Misra, “Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets,” arXiv:2201.02177 (2022). For Bengio’s System 2 program, see Yoshua Bengio, “Deep Learning for System 2 Processing,” AAAI 2020 tutorial materials, accessed May 21, 2026, <https://www.iro.umontreal.ca/~bengioy/AAAI-9feb2020.pdf>. For LeCun’s world-model program, see Yann LeCun, “A Path Towards Autonomous Machine Intelligence,” version 0.9.2, June 27, 2022, OpenReview, <https://openreview.net/pdf?id=BZ5a1r-kVsf>.

¹⁵ Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao, “ReAct: Synergizing Reasoning and Acting in Language Models,” arXiv:2210.03629 (2022); Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom, “Toolformer: Language Models Can Teach Themselves to Use Tools,” arXiv:2302.04761 (2023); Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen, “A Survey on Large Language Model Based Autonomous Agents,” arXiv:2308.11432 (2023).

¹⁶ Mahowald et al., “Dissociating Language and Thought in Large Language Models”; Bender and Koller, “Climbing towards NLU”; Bender et al., “On the Dangers of Stochastic Parrots.”

- ¹⁷ Alessio Montaruli, *Executive Search as Erfahrung: Profiles, Persons, and Professional Judgment in the Age of AI* (2027), Introduction and chap. 1. For the wider philosophical distinction, see Martin Jay, *Songs of Experience: Modern American and European Variations on a Universal Theme* (Berkeley: University of California Press, 2005); Hans-Georg Gadamer, *Truth and Method*, trans. Joel Weinsheimer and Donald G. Marshall, 2nd rev. ed. (London: Continuum, 2004), 346 to 362.
- ¹⁸ Montaruli, *Executive Search as Erfahrung*, chaps. 1, 5, and 6. The argument there distinguishes generated output, reader-side *Erlebnis*, and professional *Erfahrung* under conditions of evidence, encounter, correction, and responsibility.
- ¹⁹ Montaruli, *The Vectorized Afterlife of the They*, Introduction and chaps. 4 to 5.
- ²⁰ Montaruli, *The Vectorized Afterlife of the They*, chap. 8; Montaruli, *Executive Search as Erfahrung*, chap. 7.
- ²¹ Montaruli, *The Vectorized Afterlife of the They*, chap. 9.
- ²² Montaruli, *The Vectorized Afterlife of the They*, chaps. 6 and 9; Montaruli, *Executive Search as Erfahrung*, chap. 7.
- ²³ Montaruli, *The Vectorized Afterlife of the They*, chap. 11; Montaruli, *Executive Search as Erfahrung*, chap. 7.
- ²⁴ Montaruli, *Executive Search as Erfahrung*, chap. 7. See also Walter Benjamin, “Experience and Poverty,” in *Selected Writings, Volume 2, Part 2: 1931 to 1934*, ed. Michael W. Jennings, Howard Eiland, and Gary Smith, trans. Rodney Livingstone et al. (Cambridge, MA: Harvard University Press, 1999), 731 to 736.
- ²⁵ Carlos E. Perez, “Defining Artificial Fluent Systems,” *Intuition Machine*, accessed May 21, 2026, <https://medium.com/intuitionmachine/defining-artificial-fluent-systems-ae162014b4db>; Sequoia Capital, “AI’s \$600B Question,” accessed May 21, 2026, <https://sequoiacap.com/article/ais-600b-question/>; Goldman Sachs Global Investment Research, “Gen AI: Too Much Spend, Too Little Benefit?” *Top of Mind*, June 25, 2024.
- ²⁶ Stanford Institute for Human-Centered AI, *The 2025 AI Index Report* (Stanford, CA: Stanford HAI, 2025); MUFG Americas, “The AI Weekly: Venture Capital and Private Sector Activity,” February 27, 2026; Value Add VC, “AI Company Valuations in 2025: How the Top AI Startups Are Being Priced,” accessed May 21, 2026, <https://valueaddvc.com/blog/ai-company-valuations-in-2025-how-the-top-ai-startups-are-being-priced>.

- ²⁷ Sequoia Capital, “AI’s \$200B Question,” accessed May 21, 2026, <https://sequoiacap.com/article/follow-the-gpus-perspective/>; Sequoia Capital, “AI’s \$600B Question,” accessed May 21, 2026, <https://sequoiacap.com/article/ais-600b-question/>; Goldman Sachs Global Investment Research, “Gen AI: Too Much Spend, Too Little Benefit?” *Top of Mind*, June 25, 2024; Daron Acemoglu, “The Simple Macroeconomics of AI,” *Economic Policy* 40, no. 121 (2025): 13 to 58.
- ²⁸ Alex Engler, “Anthropomorphic AI Terms Create Gaps in Accountability,” Brookings Institution, accessed May 21, 2026, <https://www.brookings.edu/articles/anthropomorphic-ai-terms-create-gaps-in-accountability/>; Public Citizen, *Chatbots Are Not People: Designed-In Dangers of Human-Like A.I. Systems*, accessed May 21, 2026, <https://www.citizen.org/article/chatbots-are-not-people-dangerous-human-like-anthropomorphic-ai-report/>.
- ²⁹ Bender et al., “On the Dangers of Stochastic Parrots”; Drew McDermott, “Artificial Intelligence Meets Natural Stupidity,” *ACM SIGART Newsletter* 57 (1976): 4 to 9; Melanie Mitchell, “If It Quacks Like a Duck,” accessed May 21, 2026, <https://melaniemitchell.me/EssaysContent/TLS2025.pdf>.
- ³⁰ For “electronic brain” and expert-system precedents, see “World’s Smallest Electric Brain,” *Radio-Electronics* (1950), accessed May 21, 2026, <https://www.ccitalia.net/descarga/docs/1950-simon-berkeley-radio-electronics.pdf>; “AI Winter: Understanding the Cycles of AI Development,” DataCamp; “The Rise, Fall, and Resurrection of IBM Watson Health,” Garwood Center for Corporate Innovation, accessed May 21, 2026, https://corporateinnovation.berkeley.edu/wp-content/uploads/2020/04/The-Rise-Fall-and-Resurrection-of-IBM-Watson-Health_final.pdf.
- ³¹ McKinsey, “The State of AI: Global Survey 2025,” accessed May 21, 2026, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>; Boston Consulting Group, “AI Adoption Puzzle: Why Usage Is Up But Impact Is Not,” accessed May 21, 2026, <https://www.bcg.com/publications/2025/ai-adoption-puzzle-why-usage-up-impact-not>; Vistage, “Enterprise AI Adoption Is Real: Why ROI Discipline Matters,” accessed May 21, 2026, <https://www.vistage.com/research-center/business-operations/business-technology/enterprise-ai-adoption-roi/>.
- ³² Montaruli, *Executive Search as Erfahrung*, chap. 7.
- ³³ Kate Crawford and Vladan Joler, “Anatomy of an AI System,” 2018, accessed May 21, 2026, <https://anatomyof.ai/>; Engler, “Anthropomorphic AI Terms Create Gaps in Accountability”; A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen

Nissenbaum, “Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning,” in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (New York: ACM, 2022), 864 to 876.

³⁴ Engler, “Anthropomorphic AI Terms Create Gaps in Accountability”; Crawford and Joler, “Anatomy of an AI System”; Cooper, Moss, Laufer, and Nissenbaum, “Accountability in an Algorithmic Society.”

³⁵ Regulation (EU) 2024/1689, Artificial Intelligence Act; Regulation (EU) 2016/679, General Data Protection Regulation, art. 22; Court of Justice of the European Union, *OQ v. Land Hessen*, Case C-634/21, judgment of December 7, 2023; Ropes & Gray, “Helping Hand or Complete Control? AI in Recruitment in the EU and UK,” accessed May 21, 2026, <https://www.ropesgray.com/en/insights/viewpoints/102mpug/helping-hand-or-complete-control-ai-in-recruitment-in-the-eu-and-uk>.

³⁶ Montaruli, *Executive Search as Erfahrung*, chaps. 7 and 12.

³⁷ Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, “Generative AI at Work,” NBER Working Paper no. 31161 (2023); Stanford Report, “What Is AI Complementarity?” accessed May 21, 2026, <https://news.stanford.edu/stories/2025/07/explainer-ai-complementarity-llms-workplace-jobs/>; “AI, Productivity, and Labor Markets: A Review of the Empirical Evidence,” accessed May 21, 2026, <https://laweconcenter.org/resources/ai-productivity-and-labor-markets-a-review-of-the-empirical-evidence/>.

³⁸ Montaruli, *The Vectorized Afterlife of the They*, chaps. 9 and 11; Montaruli, *Executive Search as Erfahrung*, chap. 4; Bernard Stiegler, *Taking Care of Youth and the Generations*, trans. Stephen Barker (Stanford: Stanford University Press, 2010).

³⁹ Alex Capri, “Who Will Save the World from a US-China AI Arms Race?” Asia Research Institute, National University of Singapore, accessed May 21, 2026, <https://ari.nus.edu.sg/app-essay-alex-capri/>; “US-China Perspectives on Extreme AI Risks and Global Governance,” arXiv:2407.16903 (2024); “Existential Risk: The AI Safety Movement in the United States and China,” accessed May 21, 2026, <https://scholarworks.wm.edu/items/651a0752-1971-4f63-b0c5-e8e7273345eb>.

⁴⁰ Sequoia Capital, “AI’s \$600B Question”; Goldman Sachs Global Investment Research, “Gen AI: Too Much Spend, Too Little Benefit?”; Engler, “Anthropomorphic AI Terms Create Gaps in Accountability”; Crawford and Joler, “Anatomy of an AI System.”

⁴¹ Montaruli, *Executive Search as Erfahrung*, Introduction and chaps. 2 to 3; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015); Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens*

Democracy (New York: Crown, 2016); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018).

⁴² Montaruli, *Executive Search as Erfahrung*, Introduction and chaps. 8 to 11.

⁴³ Montaruli, *Executive Search as Erfahrung*, chap. 11.

⁴⁴ Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy, "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," arXiv:1906.09208 (2019); Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women," Reuters, October 10, 2018; "Automating Discrimination: AI Hiring Practices and Gender Inequality," *Cardozo Law Review*, accessed May 21, 2026, <https://www.cardozolawreview.com/automating-discrimination-ai-hiring-practices-and-gender-inequality/>.

⁴⁵ Raghavan et al., "Mitigating Bias in Algorithmic Hiring"; EPIC, "HireVue, Facing FTC Complaint from EPIC, Halts Use of Facial Recognition," accessed May 21, 2026, <https://epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/>; U.S. Equal Employment Opportunity Commission, "Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures under Title VII," May 18, 2023.

⁴⁶ Jonas Goergen, Emanuel de Bellis, and Anne-Kathrin Klesse, "AI Assessment Changes Human Behavior," *Proceedings of the National Academy of Sciences of the United States of America* 122, no. 25 (2025): e2425439122.

⁴⁷ Isabel Fernández-Mateo and Roberto M. Fernandez, "Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs," *Management Science* 62, no. 12 (2016): 3636 to 3655; Montaruli, *Executive Search as Erfahrung*, chap. 9.

⁴⁸ Boris Groysberg, *Chasing Stars: The Myth of Talent and the Portability of Performance* (Princeton: Princeton University Press, 2010); Boris Groysberg, Ashish Nanda, and Nitin Nohria, "Can They Take It with Them? The Portability of Star Knowledge Workers' Performance," *Management Science* 54, no. 7 (2008): 1213 to 1230.

⁴⁹ "M. D. Anderson Breaks With IBM Watson, Raising Questions about Artificial Intelligence in Oncology," PubMed record, accessed May 21, 2026, <https://pubmed.ncbi.nlm.nih.gov/30053147/>; Casey Ross and Ike Swetlitz, "IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show," *STAT*, 2018; "The Rise, Fall, and Resurrection of IBM Watson Health," Garwood Center for Corporate Innovation.

⁵⁰ OECD.AI Policy Observatory, “Babylon Health AI Chatbot Faces Criticism over Unsafe Medical Advice and Data Privacy Breach,” AI Incident, February 26, 2020, accessed May 21, 2026, <https://oecd.ai/en/incidents/2020-02-26-bc4b>; “Bad Bots: How Should Doctors Respond to Untested Technologies?” accessed May 21, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC6532823/>; Patient Safety Learning, “Babylon Health Lashes Out at Doctor Who Raised AI Chatbot Safety Concerns,” accessed May 21, 2026, <https://www.pslhub.org/blogs/entry/583-babylon-health-lashes-out-at-doctor-who-raised-ai-chatbot-safety-concerns/>.

⁵¹ Shruthi Shekar, Pat Pataranutaporn, Chethan Sarabu, Guillermo A. Cecchi, and Pattie Maes, “People Overtrust AI-Generated Medical Advice despite Low Accuracy,” *NEJM AI* 2, no. 6 (2025), <https://doi.org/10.1056/AIoa2300015>; “The Illusion of Readiness in Health AI,” arXiv:2509.18234 (2025).

⁵² *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023); Association of Corporate Counsel, “Practical Lessons from the Attorney AI Missteps in *Mata v. Avianca*,” accessed May 21, 2026, <https://www.acc.com/resource-library/practical-lessons-attorney-ai-missteps-mata-v-avianca>.

⁵³ Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, “Machine Bias,” ProPublica, May 23, 2016; Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, “How We Analyzed the COMPAS Recidivism Algorithm,” ProPublica, May 23, 2016; Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, “Inherent Trade-Offs in the Fair Determination of Risk Scores,” arXiv:1609.05807 (2016); Alexandra Chouldechova, “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments,” *Big Data* 5, no. 2 (2017): 153 to 163.

⁵⁴ New York State Department of Financial Services, *Report on Apple Card Investigation*, March 2021; Osonde A. Osoba and William Welser IV, “Did No One Audit the Apple Card Algorithm?” RAND Corporation, November 2019; Dan Ennis, “Goldman Cleared of Bias Claims in NYDFS’s Apple Card Probe,” *Banking Dive*, March 24, 2021. The paper’s claim is limited to explainability and accountability at the customer interface, not to a finding of broad disparate impact.

⁵⁵ *Moffatt v. Air Canada*, Civil Resolution Tribunal, 2024; McCarthy Tétrault, “Moffatt v. Air Canada: A Misrepresentation by an AI Chatbot,” accessed May 21, 2026, <https://www.mccarthy.ca/en/insights/blogs/techlex/moffatt-v-air-canada-misrepresentation-ai-chatbot>; American Bar Association, “BC Tribunal Confirms Companies Remain Liable for Information Provided by AI Chatbot,” February 2024.

⁵⁶ Office of Qualifications and Examinations Regulation, *Awarding GCSE, AS, A Level, Advanced Extension Awards and Extended Project Qualifications in Summer 2020: Interim Report* (London: Ofqual, 2020); “The Algorithm Will Screw You: Blame, Social Actors and the 2020 A Level Results Algorithm on Twitter,” accessed May 21,

2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10370707/>; upReach, *Ofqual Algorithm Research*, August 17, 2020; Richard Adams, “Ofqual’s A-Level Algorithm: Why Did It Fail to Make the Grade?” *The Guardian*, August 21, 2020.

⁵⁷ Turnitin, “AI Writing Detection Update from Turnitin’s Chief Product Officer,” accessed May 21, 2026, <https://www.turnitin.com/blog/ai-writing-detection-update-from-turnitins-chief-product-officer>; Turnitin, “Understanding the False Positive Rate for Sentences of Our AI Writing Detection Capability,” accessed May 21, 2026, <https://www.turnitin.com/blog/understanding-the-false-positive-rate-for-sentences-of-our-ai-writing-detection-capability>; Vanderbilt University Brightspace, “Guidance on AI Detection and Why We’re Disabling Turnitin’s AI Detector,” August 16, 2023; “The Challenge of AI Checkers,” University of Nebraska Center for Transformative Teaching, accessed May 21, 2026.

⁵⁸ Pasquale, *The Black Box Society*; O’Neil, *Weapons of Math Destruction*; Eubanks, *Automating Inequality*; Daniel Kahneman, Olivier Sibony, and Cass R. Sunstein, *Noise: A Flaw in Human Judgment* (New York: Little, Brown, 2021); “When Noise Mitigates Bias in Human-Algorithm Decision-Making: An Agent-Based Model,” accessed May 21, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12747352/>.

⁵⁹ For the contrast between bounded pattern recognition and judgment-class deployment, see Montaruli, *Executive Search as Erfahrung*, chaps. 2, 5, and 12. For examples of bounded pattern-recognition or transformation tasks with clearer target structures, see John Jumper et al., “Highly Accurate Protein Structure Prediction with AlphaFold,” *Nature* 596 (2021): 583 to 589; Scott Mayer McKinney et al., “International Evaluation of an AI System for Breast Cancer Screening,” *Nature* 577 (2020): 89 to 94; Papineni et al., “BLEU”; Lin, “ROUGE”; Zhang et al., “BERTScore.”

⁶⁰ Bender and Koller, “Climbing towards NLU”; Mahowald et al., “Dissociating Language and Thought in Large Language Models”; McCoy et al., “Embers of Autoregression”; Montaruli, *The Vectorized Afterlife of the They*, chap. 11; Montaruli, *Executive Search as Erfahrung*, chaps. 7 and 12.

⁶¹ Bender and Koller, “Climbing towards NLU”; Mahowald et al., “Dissociating Language and Thought in Large Language Models”; McCoy et al., “Embers of Autoregression.”

⁶² Montaruli, *The Vectorized Afterlife of the They*, Introduction and chap. 9.

⁶³ Montaruli, *The Vectorized Afterlife of the They*, chap. 11.

⁶⁴ Montaruli, *Executive Search as Erfahrung*, Introduction and chaps. 1, 5, and 7.

⁶⁵ Sequoia Capital, “AI’s \$600B Question”; Goldman Sachs Global Investment Research, “Gen AI: Too Much Spend, Too Little Benefit?”; Engler, “Anthropomorphic AI Terms Create Gaps in Accountability”; Crawford and Joler, “Anatomy of an AI System.”

⁶⁶ Montaruli, *Executive Search as Erfahrung*, chaps. 2 and 12; Brynjolfsson, Li, and Raymond, “Generative AI at Work”; McCoy et al., “Embers of Autoregression.”

⁶⁷ Raghavan et al., “Mitigating Bias in Algorithmic Hiring”; Goergen, de Bellis, and Klesse, “AI Assessment Changes Human Behavior”; *Mata v. Avianca*; Angwin et al., “Machine Bias”; New York State Department of Financial Services, *Report on Apple Card Investigation*; *Moffatt v. Air Canada*; upReach, *Ofqual Algorithm Research*; Turnitin, “AI Writing Detection Update.”

⁶⁸ Montaruli, *Executive Search as Erfahrung*, chaps. 4, 7, and 12.

⁶⁹ Montaruli, *Executive Search as Erfahrung*, chap. 4 and coda; Stiegler, *Taking Care of Youth and the Generations*.