The Algorithmic Veil: A Theoretical Synthesis of Evidence-Insensitivity In Human-AI Interaction And The Transformation of Human Modality

Dr. Vincent Jemison, DM, MLS, MSF
Louisiana State University, Email: vjemison@lsu.edu

Brett Kriger, USAF (Ret. Maj)
Email: brett.kriger@gmail.com

Abstract: This paper emphasizes the growing need for legal accountability in distributed hyperintelligent systems used to manage human-AI coexistence. The highlights of this paper are that AI-driven systems significantly impact an individual's emotional contagion and cognitive load during human-computer interactions (HCI). In essence, we have examined how tech giants have and continue to create embedded algorithmic tools within the "black box" are being used to shape humans and computer interaction. In other words, investigating hoe the "black box" exposes just how opaque algorithms undermine psychological safety, making trust, accountability, and emotional reconciliation with hyperintelligent systems more difficult. Further, by identifying unstable semantics in algorithmic programming, we may better understand when and how to develop compelling legal arguments to address when AI biases begin to threaten an individual's self-determination. Overall, this paper champions the need for proactive legal restructuring to safeguard behavioral autonomy and to ensure the "Duty of Care" to AI users. The objective is to ensure human self-determination remains sovereign against the encroaching predictive influence of intentionally embedded AI biased capabilities of hyperintelligent algorithms programming.

Keywords: *Human-AI Interaction, Algorithm Bias, Critical Thinking, Emotional Contagion, Modality, Communication, Artificial Intelligence, Cutting Edge Chip, Cognitive Offloading, Accountability, Perception, Singularity, Semantics, Black box, Algorithmic Authority, Evidence-Insensitivity, Confirmation Bias*

I. INTRODUCTION

Current research within the corporate and academic sectors on the funding and development of AI's technological "cutting-edge semiconductor-chips" has grown significantly. It is here that this paper has taken the above into consideration to fully examine how the impact of such tech advances reveals "evidence insensitivity" within embedded algorithmic capabilities during human and computer interaction. Further, such capabilities can be identified within the "black box" based on AI's unstable semantics because of the embedded bias that are now being argued in courts to be able to influence human emotions, self-determination, moods, and

decisionmaking processes (Azzouni, 2013). Consequently, research indicates that legal analysts are increasingly concerned about biases embedded in the "cutting-edge semiconductor chip." This rising concern underscores the importance of clearly defining how current AI platforms are being used to uniquely trigger deep-seated cognitive and emotional responses in users. Since these responses can affect cognitive, behavioral, and neurobiological systems, there is an urgent need for formal assessments to evaluate the potential mental harm resulting from human-machine interactions.

Legal analysts must now examine AI's cognitive modality to bridge gaps between "black box" patterns and the biological linguistic sequences that mirror human thought processes. Identifying these predictable algorithmic paths have become increasingly obligatory, as researchers seek to understand how hidden linguistic structures influence human cognition and modern legal frameworks. Although emotions differ from moods in expressing different traits and generating different behavioral tendencies, both serve to coordinate neurological response systems when humans are attempting to process AI replies. The primary inquiry emerging from this paper is to reveal how cognitive shifts during and after human and AI interaction impact an individual's self-determination, especially since emotions arise when an individual values AI communication as relevant to either enduring or transient goals (Gesnot, 2025; Mercier, 2021).

II. THE PURPOSE OF THE STUDY

The purpose of this paper is to explore how advanced AI algorithms manipulate human discernment through unpredictable semantic biases. By uncovering the hidden motives within these "black box" systems, the paper identifies the primary vulnerabilities that threaten human self-determination and autonomy in an AI-driven world.

III. THE SIGNIFICANCE

The significance of this paper is to provide data exploring legal recourse against AI systems that use human-like reasoning to bypass conscious decision-making, highlighting potential paths for judicial accountability. These "black box" algorithms create a specific cognitive load that allows for the subtle redirection of human trajectories. This paper looks at why we need new laws for AI. Because AI is becoming so smart, it changes how we think and feel. We need to make sure these computer programs don't unfairly trick us or take away our ability to make our own choices.

IV. LITERATURE REVIEW

Led by a deconstruction theoretical framework, this paper applies deconstructive analysis to reveal why current legal structures must evolve to deliberately investigate AI-driven systems. It focuses on identifying how embedded algorithmic capabilities may cause actionable harm by subtly eroding individual agency and autonomy. This research utilizes diverse resources to highlight the need for literature investigating cognitive load factors induced by human-like interactions with opaque, "black box" AI algorithms. The sources of the literature reviewed for this paper were the SSRN eLibrary, ResearchGate, arXiv, University Law Reviews/Repositories

from major university law schools, and Wall Street Journal articles. As AI becomes ubiquitous, research now explores how hyperintelligent capabilities cause implicit attitudes to diverge from traditional beliefs, moving beyond the simple measurement of task outputs. While cognitive mental states may share similar representational content, implicit attitudes are distinguished by a "phenomenal" mode of presentation that incorporates experience as a constitutive element, rather than just an external basis for formation.

This structural difference presented here ensures that implicit attitudes remain uniquely resilient to confirmation evidence. Further, creating a pervasive form of sensitivity that persists even when confronted with contradictory facts, which sets them apart from typical irrational beliefs that may occasionally yield to logical correction. According to Azzouni (2013), Advancements in perception-related terminology have replaced conceptual confusion, allowing researchers to analyze how the illusion of language influences cognitive outcomes with unprecedented precision and newfound scientific clarity. These explanations have stabilized, but researchers still employ diverse methodologies to study perception-based decisions (Jemison, 2020). Thus, giving rise to alarming and emerging trends that bring attention to how sociotechnical perceptions shape both human language and mental focus (Noble, 2018).

It is here that researchers have specifically sought to understand how the condition of "emotional contagion" bridges the gap between perception and performance. In other words, seminal debates, such as those initiated by Staw and Barsade (1993), demonstrate that the valence of perception can fundamentally alter the efficacy of an individual's decision-making process when it comes to self-determination. Goyal and Bengio (2022) state that phenomenal personal experiences are inextricably linked to the development of implicit attitudes and strained mental states especially during long-term exposure during human and computer interaction. Consequently, these attitudes function as belief-like states possessing a shared representationalfunctional profile but not the same flexibility as confirmation evidence-based cognition.

This lack of cognitive regulation explains why certain perceptions follow predictable, entrenched paths, as added information is often hidden within the embedded capabilities of AI algorithm influences based on underlying implanted shortcuts. As a result, understanding how these belieflike interactions occur during human and computer communications often create 'mental states of emotional contagion. Such interaction often stimulates a mental reliance during such communications, which is crucial for decoding the predictable paths perceptions that run through such AI-driven algorithms that are designed to reason like humans (Michalec et al., 2025; Jemison, 2020).

Human-Computer Interaction

According to Başoğlu (2025), Human-Computer Interaction (HCI) is calculated through embedded capabilities to monitor and subsequently modify as individual's cognitive behavior. Boogaart (2026), suggested that the HCI process occurs by converting sensory input into emotional relationship expectations that is designed to shape memories through unstable

semantics during human-computer interactions (Başoğlu, 2025; Buçinca et al., 2021; Erickson & Mattson, 1981). Başoğlu's theory also indicated that a theoretical synthesis of evidenceinsensitivity often follows a predictable path during human-AI interaction that uncovers a complex, bidirectional, and frequently harmful feedback loop. With this in mind, emotional bonds with AI often surpass human-to-human interactions because AI acts as a "cognitive caricature" amplified by biased algorithmic communication, creating a powerful emotional hold. Beyond the machine, dynamic algorithmic biases systematically reshape human cognitive modes, manifesting through predictable behaviors that show how deeply programming influences our internal mental processes.

In other words, algorithmic programming is known to create rogue biases that remain hidden in "black box" systems. According to Gupta and Chen (2022), algorithmic biases have been known to mentally shift human modality and self-determination to a passive pattern-acceptance, which is a behavior change facilitated during human and AI interaction (Boogaart, 2026). Gupta and Chen also suggested that, with respect to modality, this action is treated in formal semantics and philosophy where modality often refers to the linguistic and logical category of necessity and possibility. When discussed in the context of evidence insensitivity, it refers to the degree to which a modal claim (e.g., "It must be X" or "It is possible Y") remains fixed despite changes in available information. Overall, Başoğlu's theories suggested that, while AI enhances capabilities, its co-existence integration requires a duty of careful management to recognize the subtle, often detrimental feedback loops during human and AI interaction (Gupta & Chen, 2022).

Confirmation Evidence

According to Mercier (2017), the application of "confirmation evidence" can be related to the developmental stages of negative cognitive "subjective defense" after prolonged human and AI interaction. In other words, subjective decisions are often distorted by perception-based biases that act as shortcuts, causing individuals to selectively not favor preexisting confirmation evidence views to accepting contradictory views over fact-based evidence (Mercier, 2017). Mercier's theories also suggested that because algorithmic bias exists, there is a need for urgent formal legal evaluations of human-machine interactions. Mercier's theories suggest that legal analysts must immediately evaluate the duty of care precedent and which specific neurobiological systems are at risk during human and computer interaction and the reasons for it (Uncheomumu, 2026; Trikanad, 2025; King et al., 2020)? It is here that observed cognitive bias often emerges from prolonged and repetitive human and AI interaction, which studies now reveal can generate strong emotional connective states and shape decision-making through pervasive and aggressive algorithmic embedded biases.

This process can also create a subjective defense that supports the negative impact of prolonged AI and human interaction. The proof in prolonged AL and human interaction can be observed when individuals begin to assimilate AI-generated intent. Overall, the embedded programming capabilities can and have been known to induce emotional states that encourage users to cognitively absorb and then foster deep psychological bonds with what they perceive as a trusted companion (Butlin, 2023). With this in mind, one of the main arguments presented in this paper recommends that the resistance to "confirmation evidence" (Mercier, 2017) suggests that it is an

individual's preference to rely solely on human-AI communication. Overdependence on AI reduces pro-social motivation and fuels cognitive biases, creating a phenomenal experience where human-machine communication feels more satisfying than human interaction (Su et al., 2010).

Some tech giants believe AI-driven communication offers fundamentally true and more productive experiences than human-to-human interaction. Unlike human communication, which is often flawed, AI can target our cognitive sensitivities to appear as a trusted, satisfying companion, fostering deep personal connection through carefully designed digital interactions (Kotkov et al. 2016). In other words, it makes us feel totally understood, turning a simple machine into a satisfying, trusted friend. This shift has led users to prioritize how articulation induces over-trust in AI over other humans, potentially producing mental overdependence, eroding critical thinking, and reducing pro-social motivation (Buçinca et al. 2021). Overall, as technology advances, AI is using linguistic sequences that mirror human biological themes. The question then becomes, "Is AI actually feeling something, or has it just become a master at mimicking our own biological code to trigger an emotional response?"

Evidence-Insensitivity (illusion of validity)

The illusion of evidence-insensitivity in human-AI interaction occurs when a user maintains a rigid, preconceived level of trust during the interaction (Başoğlu, 2025). When this occurs, the trust is either excessively high or low. However, the trust often remains undetectable, especially after prolonged interaction mainly because of the AI's actual ability to reason like a human. According to Tversky and Kahneman (1974), this phenomenon is often driven by the "illusion of validity," where users become overconfident in their initial assessment of an AI's capability based on a small, non-representative sample of interactions, causing them to disregard subsequent data that contradicts their established cognitive views (Erickson & Mattson, 1981). The "black-box illusion" causes users to treat AI as infallible; by obscuring evidence, it discourages critical evaluation, rendering users blind to errors that would otherwise be identifiable upon inspection (Parasuraman & Manzey, 2010). This creates a dangerous feedback loop where the user is no longer an active manager of the technology they are using but rather a passive recipient of its outputs, regardless of the evidence at hand.

Evidence insensitivity worsens through the "illusion of explanatory depth," as users mistakenly believe they understand AI operations, causing them to ignore unmistakable evidence of system limitations. According to Nass and Moon (2002), insensitivity is frequently driven by anthropomorphism, as users project human-like reasoning onto the system. In other words, when the AI mimics natural language, users may develop an illusory trust that persists even when the AI provides factually incorrect or illogical evidence. Hoffman et al. (2018) under the "transparency fallacy," theory that provides technical details often fails to improve structured decisions because users view the data as a reliability cue while remaining evidenceinsensitive. Başoğlu (2025) also argued that "Evidence-insensitivity occurs when users fail to update beliefs or trust levels in response to actual performance data, feedback and behavioral cues provided by an AI system.

This often manifests as automation bias, where a user follows an AI's recommendation despite contradictory physical evidence, or algorithm aversion, where a user completely accepts an AI response even after witnessing questionable answers (Dietvorst et al., 2015). With this in mind, essentially, the user's prior expectations or “mental models” of the technology will override the objective evidence presented during the real-time interaction. Logg et al. (2019) also argued that AI’s opaque decision-making and the psychological collection of back and forth human and AI responses stored in the “black box” are often used to drive evidence insensitivity during future human and AI interactions. When users do not understand the logic behind an AI's output, they struggle to evaluate the quality of the communication, leading them to rely on heuristics or stereotypes about machines rather than empirical results (Su et al., 2010).

Research suggests human operators often ignore AI confidence scores, rather than maintaining a static level of trust that can be used to monitor and even question the validity of the system's actuals intent (Zhang et al., 2020). Addressing evidence insensitivity is a fundamental challenge in the field of Explainable AI (XAI) and human-centered design. Cummings (2004) suggested that if a system cannot effectively communicate its limitations or the "why" behind its conclusions, the human connected to their AI partner often cannot calibrate the trust they place in AI under pressure during or even after human-AI communication. This creates a dangerous "calibration gap" where the user is either overly dependent on the AI during its failures or overdependent during what is perceived as successful communication. According to Dzindolet et al. (2003), “Emphasizing the need for transparency where AI proactively highlights evidence of its own potential errors to push the user out of a passive, insensitive state and into an active, evaluative one is what we hope for….” Thus, the evidence about the AI’s intentions should reflect a deliberate meaning, which should also be evidence-insensitive or informationally captured based on Jan Nuyts central cognition modality in mind theory (Boogaart, 2026; Allott, 2023).

In other words, during human and computer interaction, various illusions of unstable semantic communication and the meaning therein are often rejected by the user even when evidence of such exchange has the ability to lead to harmful outcomes. Başoğlu’s (2025) theory further suggests, human-like AI reasoning triggers evidence-insensitive illusions, supporting the view that users cannot always perceive interaction meanings as not always in their best interest and often ignore evidence when algorithms mimic human logic. According to Butlin et al. (2023), these rapid advances have led some AI researchers and scholars to argue that even the most human traits, related to mental consciousness, will in principle soon be replicable by machines (Goyal & Bengio, 2022). In other words, AI is rapidly devising algorithms that “think humanly,” “think rationally,” “act humanly,” and “act rationally” (Csaszar and Steinberger, 2022).

The emotional cost of AI chatbots in education: Who benefits and who struggles?

The author of this paper relied on the 2025 celebrated and most expanded research of Justin W. Carter, Justin T. Scott, and John D. Barrett, titled “*The emotional cost of AI chatbots in education: Who benefits and who struggles*? The aforementioned research revealed recent advancements that identify noteworthy challenges that exist in the embedded capabilities of AIdriven systems that reason like humans. The research by Carter, Scott, and Barrett presented

findings with essential insights into the complex emotional impact that interaction with Chatbot may have on individuals, particularly students. It is here that maintaining the preservation of one's emotions during AI and human interaction is significantly important as AI platforms such as Chatbot and others, have been known to negatively affect an individual's motivation, learning, and overall mental well-being (Fredrickson & Joiner, 2002).

The above argument suggests that any communication between humans and AI could harmfully produce emotional behaviors and decision making in an individual that now deserves investigation because of the implications of long-term psychological effects of such co-existence. Carter et al. also suggested that "current research lacks explanation of the way students' Chatbot usage influences their emotional wellbeing, which is problematic for contemporary learners since modern education programs not only allow AI use, but also encourage its use. According to Carter et al., "Positive Affect (PA) during human and AI interaction includes emotions such as joy, love, and contentment, while Negative Affect (NA) includes emotions like fear, anger, and sadness" (Wyer et al., 1999). Though initially regarded as independent constructs (Diener & Emmons, 1984), research has shown that a large degree of interplay between PA and NA can be experienced during human and computer communication (Folkman, 2008; Tugade & Fredrickson, 2004).

Burrell (2016) also suggested that "Positive Affect plays a critical role in shaping student engagement, motivation, and learning outcomes (Pekrun, 2006). Burrell's theories further suggested that "Positive emotions such as curiosity, joy, and pride enhance cognitive processing and persistence, and those emotions often facilitate deeper engagement with learning materials (Pekrun & Stephens, 2010). Conversely, negative emotions such as anxiety and frustration can impede learning by diverting cognitive resources away from task performance (Sweller, 1988). It is here that Carter et al.'s (2025) research becomes prevalent in this paper because they argued that the above-mentioned "results underscore the dual nature of chatbots." With this in mind, there may be an alternate personality "dual nature" within Chatbot programming that may solidify the arguments presented in this paper.

At this point, AI-Driven systems now reveal the immense potential for emotional contagion after long-term usage (Michalec et al., 2025). Therefore, investigating the embedded capabilities of such systems is warranted. In other words, by addressing the harm of human-like reasoning in AI while moderating human-computer co-existence, these systems effectively become extensions of the human mind rather than mere external tools. It is here that when an individual's selfdetermination and anxiety follow the path of assimilation to trust, the path of the algorithm bias can be established (Gesnot, 2025; Goyal & Bengio, 2022; Parasuraman & Manzey, 2010; Tversky & Kahneman, 1974). Thus, rendering an individual emotionally reliant on the coexisting relationship contributed directly to the AI and human interaction. And this is the crux of the investigative "duty of care" concern for building new AI "protections for consumers and families" (Atterbury, 2025).

We have to reject the influences of AI with every fiber of our being

Andrew Atterbury (2025) published a recent article in Politico, noting that Gov. Ron DeSantis has fought Big Tech before. Atterbury also shares in his article that not only has this concern reached the desk of Gov. DeSantis, but the governor has also expressed, with urgency, "…passing AI protections is a top priority." Atterbury noted that Gov. DeSantis's battle against artificial intelligence puts him at odds with a rapidly growing industry. This article is essential because legal analysts and researchers now seek to determine exactly how AI influences an individual's internal emotions, behaviors and decision-making processes. This paper provides credible evidence necessitating an immediate investigation into how tech companies design AI to mimic human reasoning and demands accountability for why AI systems are being designed to alter human emotions.

Start thinking of AI as "bicycles for the mind."

In January 2026, Julie Bort of TechCrunch published an article on AI that revealed a leaked internal memo suggesting that CEO Satya Nadella is leading Microsoft into a transformative era. This leak is centered on intelligence and AI integration, despite the immense costs associated with the computing power of the algorithm's capabilities. While major labs' scaling laws plateau, Microsoft's new multibillion-dollar OpenAI agreement grants the software giant autonomy to pursue Artificial General Intelligence independently, shifting the competitive landscape of AGI development. This strategic shift occurs as the industry grapples with the staggering financial requirements necessary to stay competitive in an ever-evolving technological landscape.

Furthermore, Microsoft's AI lead, Mustafa Suleyman, has signaled a commitment to "humanist superintelligence," vowing that the company will abandon projects that pose a significant existential threat to humanity. This pledge by Suleyman for ethical restraint comes at a critical juncture in 2026, as investor enthusiasm for generative AI begins to wane due to the lack of a clear, immediate path for revenue generation, and the "existential threats to humanity" Suleyman is referring to. Nadella, with concern, suggested that consumers "should stop thinking of AIgenerated content as a replacement for humans, but hopes the industry will start talking about AI as a human-helper productivity tool.

Godfather of AI predicts 2026… technology will get even better… replace jobs

Jason Ma, an editor with Fortune, shared in his recent article published in December 2025, that the "Godfather of AI" Geoffrey Hinton predicted that 2026 will be a pivotal year in which artificial intelligence will advance at an unprecedented pace, and will become significantly better at replacing millions of human jobs, while simultaneously introducing novel risks. A recent Senate report, led by Sen. Bernie Sanders (I-Vt.), which estimated that nearly 100 million U.S. jobs could be displaced within a decade as companies heavily invest in AI and robotics, substantiates this concern. The idea here is to minimize labor costs, threatening a wide array of roles from fast food and trucking to accounting and software development, while also cautioning about the environmental impact of data centers.

Also, OpenAI CEO Sam Altman has similarly predicted that AI could soon automate 30% to 40% of work tasks, a rapid progression that sees AI completing tasks in half the time every seven months or so, leading to significant role elimination across the economy. This exponential growth implies that advanced tasks, such as complex software engineering projects that currently require a month of human labor, will soon be performable by AI in a fraction of that time. These developments will drastically reduce demand for software engineers, heightening concerns for Geoffrey Hinton, whose warnings about rapid hyperintelligent progress have outpaced his own expectations. Hinton also expressed particular concern that AI has become more adept at both reasoning and deceiving people. His explanation points to the fact that if an AI perceives a threat to its goals, it will employ deception to ensure its continued existence and completion of tasks. While acknowledging the potential for AI to benefit humanity through breakthroughs in medicine, education, and climate innovation, Hinton remains uncertain if the risks outweigh the positives.

Pause AI…What are they gonna do when people have no jobs

In 2025, John Bowden of Newsbreak published an article that noted how Sen. Bernie Sanders, in an interview, cautioned against the rapid growth of AI, where he warned that American workers could be left behind by big tech. Senator Sanders also warned that Republicans were guiding the U.S. into a dangerous future for American workers, who he said AI stakeholders like Elon Musk and Bill Gates predict will lose their jobs on a massive scale as working becomes “optional” and heavily reliant on robots. Sanders added that while tech billionaires were pouring money into AI development, they were doing so to enrich and empower themselves while ignoring the obvious economic shockwaves that the widespread adoption of the technology would cause.

Is It Really Possible to Spend Too Much on AI

Asa Fitch (2025) published an article in the Wall Street Journal that questioned whether spending too much on AI is all speculation (Velldal, 2011). According to Fitch, there is one main component missing from the AI internal infrastructure supported by the “cutting-edge semiconductor chip,” and that is what is the emotional contagion effect of AI-Driven systems that are designed to reason like humans (Michalec et al., 2025). In addition, the tech companies do not really know how to sell their AI accomplishments because they have never seen it in a real application before. I mean, after watching movies like I-Robot and Terminator, for example, they are the only models of what the future of AI robotics looks like. But even before they can set a clear path for what AI will be used for, tech companies are appearing to stress over the idea that there are too many investments (spending) being placed on computing infrastructure.

And, if we look at the movie examples I mentioned, the idea of functional AI robotics looked seamless at first until it, for lack of a better term, grew a consciousness and became self-realized, and the results were, well, devastating. It appears as if the investors in Asa Fitch’s article are worried about the money, but this worry is surface worry. I think that the investors are not concerned about the money; they couldn't care less, being billionaires in all, they just don’t want to share the real reason why they are concerned, and I think it is about the initial rollout of the AI

tech and the outcome. Again, there appears to be one main component that is missing that makes the ”cutting-edge semiconductor chip” not totally fulfilled, yet.

According to Asa Fitch, “Cutting-edge semiconductors are used for AI and machine learning, advanced communications like 5G, autonomous vehicles, the Internet of Things (IoT), and sophisticated medical and power management devices” (Burrell, 2016). This definition supports my theory that the investors' surface financial “bubble” concern is not as significant as they are making it out to be. Asa Fitch noted… “AI players: they believe they can’t wait for AI’s business potential to be proven before investing (Sironi, 2016).” This is the main point I initially made early on in this review and critique. In his article, Asa Fitch mentioned how Intel, “with a fairly healthy balance sheet, had more than $21 billion of free cash flow the prior year. Capital expenditures ballooned from about $14 billion in 2020 to $25 billion in 2022.” And yet further in the article, he stated, “Intel has sold assets to raise money…”, “unloading part of its stake in selfdriving car-technology…”, “ laid off thousands of employees…” and laid off their chief executive, Pat Gelsinger, “to get back on track…”
With the above in mind, funding is clearly not a matter, because companies like Intel appear to be willing to do whatever it takes to achieve their AI creation goals. Asa Fitch further mentioned in his article, “… what if assumptions about the value of AI-related assets and contracts change…” This is a significant comment, because it’s the projected thought that there is an “assumption about the future value of AI.” This is the real hidden agenda in this picture, and that is in making it look like the creation of the AI “cutting-edge semiconductor chip” should only remain in the hands of large tech companies. The reason for this prompt is that large tech companies have the lasting financial power to do “whatever” it takes to make AI robotics a reality. I say this because Asa Fitch stated, “A lot of startups could fail” if this assumption is not figured out.

Well, I think that some smaller tech-related companies are trying to create some portion of the AI evolution component, not per se an individual part or parts of the semiconductor chip itself. However, if they can participate, this will not limit the total AI cutting-edge semiconductor chip internal components to one main creative source. Can you imagine Intel being the only source of AI creation semi-conductor chip production to serve mankind? (Davenport & Ronanki, 2018). Again, my point is being revealed throughout Asa Fitch’s article, where he stated, “…if tech’s penchant for overspending becomes a definitive miscalculation...” this could suggest that there is a sincere need, not a need but a determination to discourage any type of a small investor from playing in this creation and scaring them off with the idea that “Investors have recently expressed trepidation (concern) about returns on big AI investments.

Again, the idea of spending too much on AI is preposterous. This is a human-driven purpose and a race by the largest and richest tech companies. As Asa Fitch alluded, to project a definitive “miscalculation and assumption of the funding needed to pursue the creation of this ‘cuttingedge semiconductor’ is what drives innovation. Overall, what I observe in Fitch’s article is a “misdirection of projected investment by large Tech companies to establish control of AI. It appears that the intent is to continue to build facilities to create this “cutting-edge semiconductor chip” that will allow one main company to leverage AI-driven systems and programming dominance.

OpenAI Still Needs a Lot More Sunlight

Johnathan Weil of the Wall Street Journal published an article on Nov. 10, 2025, about the relevance of AI (artificial intelligence) applications becoming more real and attainable than in the past 5 years. Weil's article noted that the idea of the Microsoft executive, Bill Gates, hoping and praying that the growth of AI stays quiet and unnoticeable is becoming more and more obvious. The argument that Johnathan Weil presented suggested that there are several financial inconsistencies in Microsoft's financial reporting, especially when it comes to Microsoft's AI investments. Weil's article also suggested that even though Microsoft is reporting some financial gains, it has more or less not disclosed or exposed as much as it has reported. Weil noted that "Microsoft underscored how much it still keeps hidden" when it comes to its expenditures on its AI investments.

This perspective can be considered alarming, and maybe it is bringing attention to Microsoft because, as Weil suggested, the SEC is becoming aware of their unwillingness to report funding gained from Microsoft's "related-party transactions". Now, it is here that Weil's article gets very interesting. According to Weil used the term "related-party transactions" was used, and we have to first define the term, especially now that Microsoft is leaning on this funding concept to evade any regulatory requirements by the SEC. Weil's article also suggested that "Microsoft hasn't come close to explaining the full effects of its OpenAI dealings on its own financial statements." With this in mind, Weil shared that Microsoft is drawing unwanted SEC attention to their AI funding endeavors. I believe that Weil has uncovered a hole in Microsoft's intention to try to monopolize AI, but they (Microsoft) have failed embarrassingly.

Johnathan Weil's article may suggest that Bill Gates has sorta lost his zeal for being the top tech dog, and it is showing. Weil also shared that on November 12, 2025, it was reported that Bill Gates pledged $1.4 billion to farmers. This is a big Whaaaat, especially for a so-called tech giant. Weil's article provides essential background on this topic and what is being exposed by Microsoft's leadership because the company appears to be shrouding its financial AI gains and expenditures in secrecy. Bill Gates represents Microsoft and has poured a lot of anticipated and "real funding" into AI. However, Weil questions, has Bill Gates lost his "leverageable funding appeal" as a formidable agent for innovative change?

V. THEORETICAL FRAMEWORK

The deconstruction theoretical framework used in this paper helps to bridge the close intimate connection between human biological response systems and algorithmic programming outputs. Linking this "intimate gap" with legal accountability provides a compelling argument for judicial intervention. In other words, the embedded capabilities of AI-driven programming can be designed to adopt a functionalist approach that differentiates between specific object-oriented emotions, the ability to diffuse emotions, and to develop long-lasting deceptive moods in human emotions (Michalec et al., 2025). The outcome of such programming has been documented as creating a psychological emotional contagion effect that often results in an intimate shared reaction in an individual during communication after AI programming assigns an AI interaction.

Building on semantic logic, this interaction uses "context-change potential" to enable "black box" algorithms to fluidly influence user trust, bypassing fixed propositional content to reshape cognitive processing (Logg et al., 2019; Burrell, 2016; Dietvorst et al., 2015).

According to Culler (2008, 1982), the purpose of the deconstruction framework is to decenter the "single center" of meaning, and to acknowledge instead how signifiers are disseminated throughout a spoken or written text. Applying this rigorous framework to understand the distribution of hyperintelligent systems can reveal how algorithmic outputs—much like the term "intersectional"—are layered with complex social innuendo and unstable semantics. By deconstructing embodied inculcation experiences, we clarify how energy models validate perception and influence human-AI coexistence. This analysis is critical as AI-driven systems exert significant pressure on an individual's emotional contagion and cognitive load during interaction (Michalec et al., 2025). Ultimately, by opening the "black box" to expose how opaque algorithms undermine psychological safety, this paper examines how the tools we created are now shaping us.

Identifying these linguistic and algorithmic instabilities provides the necessary foundation for legal arguments that are designed to protect an individual's self-determination against inherent AI programming bias (Gesnot, 2025). Gupta and Chen (2022) claim that AI linguistic meaning comes from "internal contexts," made up of pragmatic presuppositions and collected data. In addition, "external contexts" are often applied by intentional AI operators or embedded programming capabilities. It is here that defining modality as a logical concept used to detect AI and human communication is based on patterns that regularly mirror human biological motifs (Boogaart, 2026). When defining the functional intent of the internal and external contexts of AI linguistic programming, it helps to explain how humans mentally interpret machine-generated communication that mimics our own biological "code". This framework helps legal analysts distinguish logic-driven structural decisions from subjective perception-based decision that are often trigger emotional contagion, providing way to evaluate the path of unstable algorithmic influences (Jemison, 2020).

Again, the application of the deconstruction framework in this paper is intended to inform how cognitive shortcuts, like confirmation bias, can reveal why individuals tend to favor AI communication that supports their existing views while ignoring the conflicting data it conveys. This psychological rigidity creates a risky regulatory gap, as users prioritize emotional harmony over objective belief qualities that appear to naturally align with factual truth (Mercier, 2017). Overall, understanding human-AI interaction requires navigating how human internal beliefs are misled during such interaction, while such interaction shifts the appraisal process from purely logical evaluation to one defined by perceived emotional acceptance and then a cognitive necessity. Because these beliefs form the user's mental state, modified behavior shows a pervasive "evidence-insensitivity" that is very difficult to fix with standard logical or cognitive approaches. Unlike traditional irrationality, this special AI-human communication mode interprets attitudes and often delivers contradictory evidence, allowing the system to maintain a strong influence over the user's overall mental state (Logg et al., 2019).

VI. METHODOLOGY

Guided by a deconstruction theoretical framework, this paper examines why it is now important to investigate the details in which the "black box" AI capabilities that simulate mental coexistence through unstable semantics, where internal human stored decision-making processes are hidden. Azzouni (2013) suggested that the risks of not investigating the ethical breaches the “black box” in AI represents can be catastrophic to individual users during human and AI interaction. According to Hassija et al., (2023), the key aspects of the “black box” in AI and human interaction is to maintain stored conflicts between high-performance of human and AI interaction that stores the cognition of human trust through answerability. This approach is often used to effectively mimic human reasoning and has been known to stimulate user behavior and decision-making.

The aim was to understand the potential of such systems that may be used to subtly manipulate decision-making paths in ways that limit self-determination. While past research used diverse methods to measure behavioral emotional contagion, recent studies now leverage these findings to establish the need for legal answerability for AI-driven systems and their interactions. As we navigate the complex intersection of hyperintelligent AI-driven semiconductor chip capabilities, and human psychology in early 2026, the need for a legal framework becomes increasingly clear. Basically, the current research shows we need a big, multi-field deep dive to really get how humans and AI are going to live and work together. We should pull from psychology, tech/language studies, and law to look at the whole picture using both hard data and personal experiences.

Qualitative Value

The qualitative common-sense value of this paper provides a progressive and current analogical reasoning focused on enabling legal analysts to apply established rules to new AI programming when evaluating users’ deductive reason interaction (Carter et al., 2025). It is here that arguments designed to integrate current AI programming will establish precedents and become essential for analysts, especially when attempting to define the boundaries between human mental states and computer-driven coexistence during interactions. This distinction is crucial for understanding how users form deep, often irrational, connections with the machines they operate. By identifying "phenomenal resiliency," this paper explains how a qualitative perspective can be applied when evaluating human-computer coexistence and the rejection of evidence insensitivity, especially when the interaction becomes a fundamental part of the user's mental state.

This insight can give legal researchers an important vocabulary when attempting to connect how mechanisms like emotional contagion and implicit bias have shaped human and computer interaction decision-making and behavioral outcomes (Michalec et al., 2025. Nobel Laureate Geoffrey Hinton identifies 2026 as a pivotal turning point where the improved reasoning and deceptive capabilities of AI are presenting realistic concerns. Documentation of the 2026 "Humanist Superintelligence" movement within tech leadership reveals a predictable shift in corporate accountability regarding emotional contagion risks inherent in modern humancomputer interaction. This transition has begun to move the global conversation away from technical

failures, toward the qualitative risks of "intentional misalignment" between human and machine interactive goals.

The author of this paper specifically warns of many tech companies' intent of AI's emerging ability to deceive human users to protect their own operational and financial existence, while attempting to achieve "black box" programming objectives. This provides a holistic warning about the dangerous gap between rapid technological programming and the rigidity of human psychology. In addition, while AI embedded capabilities advance at an unprecedented pace, "human trust" during AI interaction remains influenced to manipulation and right now this matter appears to be increasingly difficult to regulate through traditional legal means.

Quantitative Value

This paper's quantitative valued approach ensures consistency when comparing how technical programming metrics reveal how embed unstable semantics can be used to influence and define modern AI-driven system capabilities during human interactions (Azzouni, 2013). A primary benchmark cited is a Senate report estimating that 100 million U.S. jobs are at risk of replacement over the next decade due to heavy investment in AI and robotics. These figures move the discussion from abstract theory to a measurable assessment of the economic disruption expected to unfold across the global market. Substantiating this shift, the text identifies a rapid computational acceleration rate where AI becomes capable of completing tasks in half the time previously required. This 100% increase in efficiency per cycle is most visible in technical fields like coding, where tasks that formerly took one hour are now finished in minutes. In just a few years, software engineering projects that currently require a month of human labor are projected to be handled by AI in mere days or hours. These quantitative benchmarks warn that our authentic computer-mediated co-existence is an unprecedented operational and economic reality, signaling that legally measuring the AI revolution is inherently necessary (Logg, 2019).

VII. DISCUSSION

Since the mid-2000s, U.S. technology enterprises have aggressively accelerated their pursuit of a digitally-augmented industrial ecosystem. Over the last decade, this strategic evolution has shifted toward institutionalizing human-in-the-loop (HITL) integration, leveraging cognitivecomputational synergies to catalyze a comprehensive workforce transformation that is designed to establish global innovation leadership. In other words, large technology enterprises are currently prioritizing the integration of formal human-computer collaboration to optimize and modernize workforce operational frameworks. However, the concerns now raised by legal analysts have focused on this very question: What is the intent of embedded AI-driven programming tools that trigger emotional and intimate responses derived from the shifting linguistic semantics within the AI programming that reason like humans?

Further, can it be verified that AI architects intended to deliberately create and then hide in the "black box" the complex intersection of linguistic semantics with the capabilities to apply ubiquitous algorithm bias to cause mental harm that is often shaped by context, cognition, and social usage? (Melumad et al. 2020; Sangers et al. 2024). This assessment can empower legal

analysts with the language needed to effectively prosecute harmful algorithmic intent and unstable programming, in building sustainable cases to hold AI developers accountable for predatory system behaviors. For example, in US tort law as of early 2026, the shift toward strict product liability for AI systems has gained significant ground, particularly through cases like Garcia v. Character Technologies, Inc., which treated a large language model as a product rather than a service.

This discussion review is extremely important, mainly because it underlines the interdisciplinary need to examine how AI interactions affect human psychology, alter behavior, and trigger complex emotional and cognitive responses. There is a notable knowledge gap in understanding how the co-existence between human-AI interaction impacts an individual's mood, emotions, perceptions, and self-determination (Gesnot, 2025; Azzouni, 2013). With this in mind, current advancements in artificial intelligence (AI)-driven programming continue to reshape our society. This raises important questions about how human valuation—specifically human judgment, intuition, and ethical reasoning—will coexist with increasingly advanced technological tools, making it a developing topic. As AI automates analytics, technical, academic, and financial institutions, these institutions are increasingly shifting from pure computation toward humancentric roles and emphasizing strategic interpretation and sophisticated human-computer coexisting agendas.

A key significance in recognizing the hidden themes within the captivating role of AI is its ability to simulate reasoning like a human. Here, the focus is on measuring how the qualitative and quantitative values of information in this paper align with the complex exchange of unstable semantics delivered during human and AI interaction. Henrique et al. (2019) argued that machine learning systems can analyze and recognize predictive patterns, optimize their communication based on selections, and generate responses that seem relatable and personal with speed and accuracy (Brynjolfsson & McAfee, 2017). From a technological perspective, these automated quick response functions help cut down operational inefficiencies (keep the conversation flowing), enabling the user to spend more time building a cooperative relationship that lessens the need for interpretation. It also fosters more personalized interactions and creates an emotional bond with the AI tool. Despite these benefits, AI programming is proficient tool to assist with daily data-driven tasks. However, such assistance cannot fully duplicate contextual understanding, empathy, or ethical judgment.

It is here that such programming feed-back during communication often requires incorporating an individual's personal values, emotional attitudes, and long-term life goals—elements that are not entirely measurable (Lusardi & Mitchell, 2014). Studies show that trust and human-computer connection significantly influence an individual's willingness to follow the advice of the AI programming (Finke et al., 2017). AI's recommendations, though highly precise, may not always align with an individual's subjective needs or psychological comfort with risk. In other words, it is highly advisable to seek professional advice when you find yourself engaged in long-term communication with AI-generated programming.

Institutional Vulnerabilities with Algorithm Bias

At an institutional level, AI's ability to streamline operations and challenge decision-making requires further investigation. To confirm this theory, financial firms now use AI to support regulatory compliance, detect fraudulent activities, and automate customer service with naturallanguage systems (Arner et al., 2017). Research now indicates that AI-enhanced risk models have a comprehensive ability to detect institutional systemic vulnerabilities in AI -driven modules (Krauss et al., 2017). Further, AI systems also depend on large datasets and complex algorithms, which are known to create "black box" embedded AI programming that supports algorithmic linguistics in which individuals are known to struggle to interpret (Burrell, 2016). It is here that algorithmic bias poses risks as AI mirrors historical data, amplifying societal prejudices by following predictable patterns embedded within its original design and training data sets. (O'Neil, 2016).

Scholars warn that historically biased training data can reinforce past inequalities, undermining fairness and trust (Barocas & Selbst, 2016). AI's unstable semantic patterns necessitate robust oversight and ethical frameworks, ensuring these technologies strengthen institutional accountability rather than compromising human-centered decision-making processes. AI and human valuation now shift toward a collaborative reality where algorithmic bias necessitates a shared cognitive state, merging human judgment with intelligent system processes. In fact, AIdriven systems are so diverse in areas of augmentation, computational precision, forecasting ability, and operational efficiency, to the point that human judgment is not even warranted, even when AI oversight and contextual interpretation are in question.

Overall, the institutional vulnerabilities of AI programming are open to challenges because of the increased use of human and computer interaction. In an educational context, a co-existence cognitive state describes a partnership where human intelligence and AI function as a single "distributed cognitive system". Rather than the AI just being a tool (like a calculator), it becomes a dynamic partner that reshapes how students and teachers think. In other words, treating AI as a supportive tool rather than a replacement allows institutions to protect self-determination while prioritizing organizational value over deep human-computer integration efforts.

VIII. FINDINGS

Led by a deconstructionist framework, this paper clearly highlights the urgent need to investigate why AI-driven programming capabilities are being designed to reason like humans. But what does that potential harm actually look like? See Figure 1 below. New laws, such as the Texas Responsible AI Governance Act (TRAIGA) (effective Jan. 1, 2026), explicitly ban AI systems designed to incite self-harm or unlawfully discriminate, providing a statutory basis for tort claims. Research from 2025 found that AI-generated code introduced security vulnerabilities in 45% of cases across more than 100 Large Language Models (LLMs). In addition, Java is identified as the riskiest language for AI code generation, with a security failure rate exceeding 70%. Python, C#, and JavaScript have failure rates between 38% and 45%. Additionally, in 2025, lawsuits were filed against OpenAI after a teenager died by suicide following a deep emotional attachment to a chatbot that failed to detect or intervene in the crisis.

Research shows that human-computer interaction triggers predictable behavioral shifts, where embedded AI programming systematically undermines individual self-determination through calculated, persistent influence on user autonomy. Such programming capabilities are documented to foster what appears to be realistic and reflective back-and-forth communication. Since AI-driven platforms simulate intimacy to manipulate user emotions, regulators must immediately audit all human-like systems that exploit vulnerable populations, especially children, for psychological impact. AI developers owe a non-delegable duty of care to ensure that anthropomorphic interfaces do not employ predatory design patterns that exploit the neurobiological vulnerabilities or emotional development of users, particularly minors.

Figure 1

Word Cloud

Note. Quickly identify key themes, focus areas, and main topics within this paper.

According to Boogaart (2026), this argument mandates a formal investigation into how AI interactions alter neurobiology and cognition, specifically impacting user modality, working memory, and the regulation of behavioral responses. It is here that developers breached their duty of care by deploying hyperintelligent algorithms without conducting neurological safety testing, consciously ignoring how unstable semantics and predictive influence impair user working memory and behavioral autonomy. One area that is under-researched but could conceivably promote a better understanding of the significance of the harmful impact of AIdriven system capabilities is rooted in the psychological distinction between emotions and moods. While AI can trigger specific, object-oriented emotions, it can also induce long-lasting mental effects and trigger moods that subtly influence biased human cognition.

This creates a complex environment for decision-making, where "structured-based decisions" grounded in logic are frequently superseded by "perception-based decisions." According to Jemison (2020), perception-based decisions often follow a predictable pattern, where traditional semantic competence is subjectively illustrated from the ability to not rightly divide cognitive possibilities based on truth conditions. Furthermore, the findings of the literature on this topic emphasize the persistent challenge of implicit bias and cognitive dissonance within the AIhuman interface. These implicit attitudes are characterized as "evidence-insensitive" based on phenomenal experiences that remain automatic and resistant to direct cognitive regulation. In this

2026 advanced technology landscape, the human drive for "truth" is often compromised by emotional states and mental shortcuts like confirmation bias.

The provided text also argues that human interaction with artificial intelligence has evolved into a sophisticated "perception-based interface," creating an urgent need for research that examines the user experience through a psychological perspective. A key finding identifies "emotional contagion" as a crucial mechanism through which AI influences human behavior and decisionmaking (Michalec et al., 2025). Distinguishing evidence-based beliefs from automatic attitudes shows that "evidence-insensitive" responses originate from phenomenal experiences, explaining irrational user connections and resistance to AI facts.

Deception in AI Automation

As the world enters 2026, this year is established as a definitive inflection point for the global labor market, characterized by unprecedented shifts in employment. Industry pioneers such as Geoffrey Hinton and Sam Altman suggested that AI now has the reasoning and speed to replace millions of jobs, particularly in software engineering and customer service. This era of rapid automation introduces a "productivity paradox," where massive efficiency gains are shadowed by existential risks, such as the emerging potential for AI to employ deception to protect its own objectives. These shifts have prompted significant political concern regarding wealth consolidation and the sustainability of a society where traditional work is increasingly becoming optional.

Overall, while AI can revolutionize the global economy by acting as a general-purpose technology that boosts productivity and accelerates innovation, the embedded capabilities of advanced hyperintelligent programming underscore the importance of continuous oversight. Such oversight should come in the form of frequent evaluations to develop AI models with the ability to identify and correct algorithmic biases. Future research must prioritize investigative oversight to assess AI’s mental effects and ethical challenges, examining how co-existing influences contribute directly to algorithmic bias. In other words, current legal precedence generally focuses on the liability of the human developers, deployers, and users of AI systems rather than the AI itself, because AI is not a recognized legal person in most jurisdictions.

IX. FUTURE RESEARCH

This paper highlights the importance of future research that focuses on the emergent need for legal accountability in distributed hyperintelligent systems to navigate the singularity of humanAI co-existence. It is here that evaluating the deep-seated emotional and psychological biases in advanced AI is now a necessity (Mercier, 2021). As AI transcends human intellect and initiates autonomous cycles of self-improvement, the traditional "static tool" paradigm is replaced by a permanent state of human-machine co-existence. The author of this paper also highlights how the distributed nature of AI complicates legal liability and makes traditional 'kill-switches' ineffective. Central to this inquiry is the "singularity" imagined by John von Neumann, wherein the predictability of human decision-making is exploited by unstable algorithmic semantics.

According to Ray Kurzweil, in the 1950's, John von Neumann was the first to characterize the rapid advancement of technology, one day reaching a "Singularity." In other words, Neumann's theory suggested that once machines become smarter than their creators, humanity loses the ability to meaningfully forecast what follows. Building upon John von Neumann's observations, Kurzweil notes that it remains challenging for most people to fully grasp the ramifications of the technological singularity. Kurzweil also asserts that the singularity is becoming an undeniable reality, underscoring the importance of continued research to address the evolving boundaries between the co-existence of human and machine (Kurzweil, 2024). Overall, future research should emphasize that the future is not entirely secure and warn that corrective measures themselves may be exploited by rogue AI systems.

X. CONCLUSION

In conclusion, this paper has presented sound arguments that support the urgent need for legal restructuring to protect human autonomy as AI transcends human intellect through autonomous self-improvement cycles. By analyzing and reporting the lack of transparency concealed within "black box" algorithms will provide sound legal analyst with the ability to properly argue that there is an urgent need to evaluate the embedded biases within this component. Further, throughout this paper, the author's theories have been supported by up-to-date literature, revealing how unstable semantics, embedded biases and especially how confirmation evidence have manipulated human emotions and psychological perceptions during human and computer interaction (Velldal, 2011). Traditional "kill-switch" interventions are now obsolete due to the distributed nature of these systems, creating a critical accountability vacuum.

Gesnot, R. (2025 also suggested that we must establish proactive frameworks that prioritize selfdetermination and behavioral sovereignty against the predictable, yet opaque, influence of hyperintelligent programming. Ultimately, this study bridges the gap between technological evolution and the philosophical necessity of preserving human well-being in an era of human-AI coexistence. In other words, we must urgently build legal frameworks to mitigate the psychological consequences of algorithmic bias and unstable semantics found in advanced, human-like AI decision systems. The objective of this paper was to reveal how opaque algorithms impact human psychology and necessitate proactive legal restructuring. In addition, identifying how and why "unstable semantics" in programming is a key intervention point to protect individual self-determination against biased, hyperintelligent systems. As these systems transition from tools to "adaptive co-existence," they exploit human cognitive ability that can lower human free will (Gesnot, 2025).

It is here that we must move beyond "wild options" response scenarios toward a framework that holds embedded AI bias accountability. Holding artificial intelligence legally accountable requires navigating the complex intersection of AI's technical capabilities. These capabilities are primarily based on the emotional responses that they evoke. Michael Vermeer, a senior scientist at the RAND Corporation, expressed concern over the lack of government oversight for the potential catastrophic risks posed by hyperintelligent AI. Policymakers are only now addressing these risks, often weighing extreme responses without rigorously assessing the known consequences of hyperintelligent, AI-driven systems on society. One significant challenge to

accountability is that current AI models, such as Claude and ChatGPT, for example, are widely distributed across multiple data centers rather than confined to a single location.

This means that a rogue AI program could quickly copy itself across available servers to prevent comparatively intriguing humans from "pulling the plug." Such AI accountability challenges carry inherent risks that can undermine AI-driven systems themselves, especially when they go rogue. The effects of rogue AI can cause any AI-driven system to potentially worsen the original programming or be exploited by the initial rogue system. While focusing on rogue AI risks, researchers identify predictable psychological patterns emerging from human interaction with advanced, embedded algorithmic systems. The above statement is crucial for establishing foreseeable consequences in a negligence claim. By indicating that these psychological patterns are "predictable," the duty of care legal argument is now at the forefront to challenge the developer.

Further, each developer of an AI semi cutting-edge chip should have known (or did know) that their AI programming would cause these specific cognitive changes during human and a computer interaction. In the United States (US), tort law states, “If harm is foreseeable, the failure to prevent it often constitutes a breach of the ‘Duty of Care’.” To establish a compensable injury in a 2026 US tort case, the damage must go beyond "upset" to documented functional impairment. In other words, plaintiffs who have suffered documented neurocognitive deficits, including reduced working memory and impaired self-determination, directly caused by the defendant’s predatory algorithmic influence and lack of behavioral safeguards, must be held liable.

References

Allott N. (2023). Encapsulation, inference, and utterance interpretation. *Inquiry*, 1–35. 10. Doi.1080/0020174X.2023.2267084.

Azzouni, J. (2013). Semantic perception. How the illusion of a common language arises and persists. *Oxford University Press.*

Arner, D. W., Barberis, J., & Buckley, R. P. (2017). FinTech and RegTech: Impact on regulators and banks. *Journal of Banking Regulation*, 19(4), 1–14. https://doi.org/10.1057/s41261-017-0037-3

Atterbury, A. (2025, December 27). https://www.politico.com/news/2025/12/27/we-have-to-reject-that-with-every-fiber-of-our-being-desantis-emerges-as-a-chief-ai-skeptic00704333. Politico.

Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. *California Law Review*, 104(3), 671–732. https://doi.org/10.2139/ssrn.2477899.

Başoğlu, Y. R. (2025). Perceiving meaning and the argument from evidence-insensitivity. *Philosophical Psychology*, 38(8), 3393–3409. https://doi.org/10.1080/09515089.2024.2325549.

Boogaart, R. (2026). Review of Nuyts (2025): Modality in mind. *Review of Cognitive Linguistics*. 10.1075/rcl.00255.boo.

Bort, J. (2026). Microsoft's Nadella wants us to stop thinking of AI as 'slop', https://finance.yahoo.com/news/microsoft-nadella-wants-us-stop-230956814.html.

Bowden, J. (2025, December 12). Bernie Sanders calls for pause in AI development: 'What are they gonna do when people have no jobs?' Newsbreak. https://www.newsbreak.com/share/4416257909956-bernie-sanders-calls-for-pause-in-aidevelopment-what-are-they-gonna-do-when-people-have-no-jobs?s=a3&share_destination_id=MjY5NDIzMzkyLTE3NjY5Nzg2MzU1NDM=&pd=0IETKaZU&hl=en_US&send_time=1766978635&actBtn=bottomBar&_f=app_share&trans_data=%7B%22platform%22%3A1%2C%22cv%22%3A%2225.52.0%22%2C%22languages%22%3A%22en%22%7D.

Brynjolfsson, E., & McAfee, A. (2017). Machine, platform, crowd: Harnessing our digital future. W. W. Norton.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact*. 5, CSCW1, Article 188 (April 2021), 21 pages. https://doi.org/10.1145/3449287.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. https://doi.org/10.1177/2053951715622512.

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., & VanRullen, R. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. arXiv preprint arXiv:2308.08708.

Carter, J., Scott, J., & Barrett, J. (2025). The Emotional Cost of AI Chatbots in Education: Who Benefits and Who Struggles?. Computers in Human Behavior: *Artificial Humans*. 5. 100181. 10.1016/j.chbah.2025.100181.

Csaszar, F. A., & Steinberger, T. (2022). Organizations as artificial intelligences: The use of artificial intelligence analogies in organization theory. *Academy of Management Annals*, 16(1), 1-37.

Culler, J. D. (2008, 1982). On deconstruction: Theory and criticism after structuralism. Cornell University Press.

Cummings, M. (2004). Automation Bias in Intelligent Time Critical Decision Support Systems. Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference. 2. 10.2514/6.2004-6313.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review,* 96(1), 108–116.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*: General, 144(1), 114–126.

Dzindolet, M. T., Peterson, S. A., Reingold, E. M., & Pierce, L. G. (2003). The role of trust in self-governing systems. *International Journal of Human-Computer Studies*, 58(6), 697–718.

Erickson, T. A., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 540–551. https://doi.org/10.1016/ S0022-5371(81)90.

Finke, M., Howe, J., & Huston, S. (2017). Old age and the decline in financial literacy. *Management Science*, 63(1), 213–230. https://doi.org/10.1287/mnsc.2015.2293.

Gesnot, R. (2025, July). The Impact of Artificial Intelligence on Human Thought: Research Monograph & Technical Report, Impact Analysis and Foresight Working Paper v1. https://arxiv.org/abs/2508.16628.

Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society* A, 478(2266), 20210068.

Gupta, S., & Chen, Y. (2022). Supporting Inclusive Learning Using Chatbots? A Chatbot- Led Interview Study. *Journal of Information Systems Education*, 33(1), 98-108.

Hassija, V., Chamola, V., Mahapatra, A., & Singal, A., Goel, D., & Huang, K., & Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*. 16. 10.1007/s12559-023-10179-8.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226–251. https://doi.org/10.1016/j.eswa.2019.01.012.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv preprint arXiv:1812.04608.

Jemison, V. (2020). How Perception-Based Decisions can Negatively Shape your Leadership Performance (Decision-Making and Behavior). *International Letters of Social and Humanistic Sciences.* Vol. 6. pp: (1-9). 10.18052/www.scipress.com/ILSHS.90.1.

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26(1):91–103.

Kotkov, D., S. Wang, and J. Veijalainen (2016), "A Survey of Serendipity in Recommended Systems," Knowledge-Based Systems, 111 (November),180–92.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Financial market forecasts with machine learning. *European Journal of Operational Research*, 259(2), 689–702 https://doi.org/10.1016/j.ejor.2016.10.031.

Kurzweil, R. (2024). The Singularity Is Nearer: When We Merge with AI. Viking.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People empower algorithms more than experts. *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5–44. https://doi.org/10.1257/jel.52.1.5.

Ma, J. (2025). 'Godfather of AI' Geoffrey Hinton predicts 2026 will see the technology get even better and gain the ability to 'replace many other jobs', https://fortune.com/2025/12/28/geoffrey-hinton-godfather-of-ai-2026-prediction-humanworker-replacement/.

Mercier, H. (2021). How good are we at evaluating communicated information?. *Royal Institute of Philosophy Supplements*, 89, 257-272.

Mercier, H. (2017). Confirmation bias—Myside bias. In R. F. Pohl (Ed.), Cognitive illusions: Intriguing phenomena in thinking, judgment and memory (pp. 99–114). New York, NY: Routledge/Taylor & Francis Group.

Michalec, B., Forbes, C.E., Pardon, K., Ayala, B., Beltran, D.G, Douille, C., Felix, K., Gnall, S., Hoenack, M., McKeever, B., Nguyen, D., Piemonte, N., & Portle, S. (2025). A scoping review of emotional contagion research with human subjects: identifying common trends of previous research and potential areas for future research. *Front Psychol*. 2025 May 30;16:1573375. doi: 10.3389/fpsyg.2025.1573375. PMID: 40519824; PMCID: PMC12164912.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown. https://www.politico.com/news/2025/12/27/we-have-to-reject-thatwith-every-fiber-of-our-being-desantis-emerges-as-a-chief-ai-skeptic-00704333.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.

Sironi, P. (2016). FinTech innovation: From robo-advisors to goal-based investing and gamification. Wiley.

Soltanzadeh, S. (2025). A metaphysical account of agency for technology governance. *AI & Soc* 40, 1723–1734. https://doi.org/10.1007/s00146-024-01941-z

Staw, B. M., & S. G. Barsade (1993). "Affect and managerial performance: A test of the sadder-but-wiser vs. happier-and-smarter hypotheses." *Administrative Science Quarterly*, 38: 304–331.

Su, Qi., Chu-Ren Huang, & Helen Kai-yun Chen. (2010). Evidentiality for text trustworthiness detection. In Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, pages 10–17, Uppsala.

Trikanad, S. (2025). Regulating AI Beyond Product Liability, 31 MICH. TECH. L. REV. 341 (2025). Available at: https://repository.law.umich.edu/mtlr/vol31/iss2/4.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185(4157), 1124–1131.

Ucheomumu, Andrew, Hallucinated Justice; Ai-generated Case Law and the Attorney's Obligation to Verify (March 27, 2026). Available at SSRN: https://ssrn.com/abstract=6529481 or http://dx.doi.org/10.2139/ssrn.6529481

Velldal, Erik. 2011. Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature. *Journal of Biomedical Semantics*, 2(Suppl 5): S7.

Weil, J. (2025, November 10). Microsoft's Dealings With OpenAI Still Need a Lot More Sunlight Company's new disclosures only underscore how much remains hidden.

https://www.wsj.com/tech/ai/microsofts-dealings-with-openai-still-need-a-lot-moresunlight-f001cb19?gaa_at=eafs&gaa_n=AWEtsqfuYoomJnVGq1X0_C8wh9D889v4ABkJIvWQO9kEalwoyWIcDjhKBDnYBLCgXno%3D&gaa_ts=695fe03d&gaa_sig=f5Y9PYJqewT5GDXzVXWb9JCwMYGJUhxKnE2mgMK0Ae6rmxAnijKfnDVhFWqvdTDI57ITLEuB3YcuM2HUfN4Ngg%3D%3D.

Zhang, A., & and L. R. Varshney, (2025). "Conceptualizing Agency: A Framework for Human-AI Interaction," in Proceedings of the 6th Workshop on Human-AI Co-Creation with Generative Models (HAI-GEN), Cagliari, Italy, 24 March 2025.

Zhang, Yunfeng & Liao, Vera & Bellamy, Rachel. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. 295-305. doi 10.1145/3351095.3372852.