

# VET: A Framework for Analyzing AI Discourse

Meredith Ringel Morris

The Paul G. Allen School of Computer Science & Engineering and The Information School  
University of Washington  
Seattle, WA, USA

merrie@cs.washington.edu

## Abstract

Public discourse on AI has become polarized; exaggerated positions on AI in traditional and social media threaten the development of AI Literacy among the general public. In this article, I introduce the VET Framework, a method for categorizing AI discourse along the dimensions of valence, effectiveness, and trajectory. I show how this framework can be used to identify, compare, and critique prevalent narratives of AI Hype, AI Doom, AI Denial, and AI Normalcy. Using VET, I analyze how each of these four stances exaggerates some aspects of the current state and/or likely evolution of AI, and illustrate how the VET framework can serve as an AI Literacy tool by supporting the “vetting” of polarized AI discourse.

## 1 Introduction

AI Literacy (Miao et al., 2024; Gunder, 2025; Long & Magerko, 2020; Tadimalla & Maher, 2025) refers to age- and role-appropriate understanding of what modern AI is (*i.e.*, what products and processes use AI), how AI works (*i.e.*, what AI is capable of and what its limitations are), and how AI is changing (*i.e.*, what is the likely trajectory for AI adoption and how will it impact society). As such, AI Literacy is a critical skill for the general public and for policymakers, in order to support appropriate and effective individual, community, corporate, and governmental uses of AI and preparation for societal changes that may result from widespread adoption of emerging AI technologies.

Formal AI Literacy curricula may play a role in educating target populations (*e.g.*, school-age students, policymakers, workers in job up-skilling or re-training courses). However, traditional media and social media discourse play a key role in shaping knowledge of AI among the general adult population, the large majority of whom do not currently have access to formal AI Literacy content (OECD, 2025; Laupichler et al., 2022).

At the time of writing (early 2026), both traditional news media and social media in the United States present polarized caricatures of both the current state of AI and its future evolution (Gilardi et al., 2024); such depictions either exaggerate or downplay the capabilities, benefits, risks, and likely development and adoption trajectory of this emerging class of technologies. While some of the narratives I discuss are present beyond the U.S., AI discourse has high global variance. For example, attitudes toward AI in China and in parts of the Global South are quite different than in the U.S. (Maslej et al., 2025). Attitudes in Europe also differ; for example, Europe’s interest in AI sovereignty (having models and data centers

that are owned by and located within EU countries) (Mügge, 2024) is quite different than current anti-datacenter backlash in the United States (Gramlich et al., 2026). Various factors contribute to these regional differences, including government messaging, regulatory frameworks, economic conditions, and political circumstances.

Exaggerated depictions of AI are an impediment to AI Literacy. I propose the VET framework for understanding the current nature of prevalent depictions of AI along the dimensions of *valence*, *effectiveness*, and *trajectory*. Using this framework, I show how different points in this taxonomy represent four common AI media narratives in the U.S.: Hype, Doom, Denial, and Normalcy. Using the VET Framework I compare and contrast these narratives and analyze their strengths and weaknesses, illustrating the framework's utility as an AI Literacy tool to "vet" the plausibility of opinions about AI.

## 2 VET: A Taxonomy for AI Discourse

To better understand current media discourse around AI in the United States, I introduce that VET Framework, which considers three facets of beliefs about AI: valence, effectiveness, and trajectory.

### 2.1 Valence

The first dimension helpful for analyzing discourse on AI is the *valence* in attitude toward AI that an author, media organization, social media post, and/or journalistic artifact exhibits. This axis ranges from ascribing extremely negative properties to AI to extremely positive ones.

Valence can relate to perceived or anticipated impacts of AI. Negative valence is often associated with concerns about impacts such as environmental externalities (Li et al., 2025; International Energy Agency, 2025), job loss (Briggs & Kodnani, 2023), impact on human relationships (Fang et al., 2025; Phang et al., 2025), the potential for differential impacts on marginalized populations (Gallegos et al., 2024), or deskilling (Kosmyrna et al., 2025; Gerlich, 2025). Positive valence tends to be associated with excitement about impacts such as elimination of tedious work (Noy & Zhang, 2023), skill expansion through human-AI teaming (Brynjolfsson et al., 2025b), augmented abilities for people with disabilities (Goodman et al., 2024; Paglialunga & Melogno, 2025), novel social support opportunities (Morris & Brubaker, 2025; De Freitas et al., 2025), and acceleration of scientific and medical discovery (Jumper et al., 2021; Bai et al., 2023).

Valence can also relate to meta-issues associated with AI, such as attitudes toward technology, privacy, capitalism and wealth inequality, and other political attitudes (*e.g.*, NIMBY-ism, Trumpism (Axios & The Harris Poll, 2026), pro- or anti-military attitudes, perceptions of China, etc.).

Finally, valence can relate to positive or negative motivations that some may ascribe to the AI itself, *i.e.*, the (controversial) belief that AI may currently or in the future be sentient (Bostrom, 2014; Yudkowsky & Soares, 2025) and may inherently have positive or negative desires as regards human society (*i.e.*, AI may or may not be aligned with human values (Christian, 2020)).

### 2.2 Effectiveness

A second useful dimension to consider is the attitude toward the *effectiveness* of AI. This dimension ranges from views that consider AI to be a weak and ineffective technology to views that consider AI to be extremely powerful and capable. Effectiveness encompasses considerations of AI's accuracy (*i.e.*, the rate and severity of errors) and reliability (*i.e.*, whether AI performs consistently and has calibration for refusing tasks it cannot perform well) (Rabanser et al., 2026). Note that determinations around

accuracy and reliability, while sometimes presented in absolute terms, may be more actionable when contextualized in relative terms, *i.e.*, with respect to a particular baseline, often typical or expert human performance for the same task (Morris et al., 2026).

Effectiveness can also reflect beliefs about direct comparisons of AI capabilities to human ones, with low effectiveness ratings reflecting the view that AI is not nearly as “intelligent” as humans (Bender et al., 2021) and higher effectiveness ratings reflecting the view that AI is approaching milestones such as AGI (Artificial General Intelligence) (Morris et al., 2025).

### 2.3 Trajectory

Finally, the *trajectory* dimension reflects attitudes about the rate of AI progress. The trajectory dimension includes attitudes about the *technical trajectory* of AI – beliefs about whether technical progress toward milestones such as AGI, robotics with human-level dexterity, level five self-driving vehicles, etc., will be fast or slow. The difference between a predicted “slow” versus “fast” technical trajectory might be quite large, since this may mean a difference in view on whether the trajectory is likely to be linear or exponential (Kaplan et al., 2020; Lu et al., 2026).

Additionally, this dimension reflects beliefs about the *adoption trajectory* of AI, such as whether individuals, corporations, and governments will incorporate AI into their lives and workflows quickly, or will be delayed by learning curves, regulation, social resistance, or other factors (Narayanan & Kapoor, 2025; Acemoglu, 2025; Brynjolfsson et al., 2021).

Finally, this dimension also reflects beliefs about the *impact trajectory* of AI. The slow end of the impact trajectory scale represents concerns such as Amdahl’s Law (Amdahl, 1967), which suggests certain aspects of tasks and processes may act as bottlenecks that limit the speed with which AI can have impact (*e.g.*, while AI may speed up parts of the drug discovery process such as identifying candidate pharmaceuticals, other aspects such as patient trials and the FDA approval process may remain slow and limit the effective speed of end-to-end AI-powered drug discovery pipelines (Lenarczyk et al., 2025)). The fast end of the impact trajectory scale predicts near-term, pervasive impacts of AI on many aspects of life (Amodei, 2024).

## 3 Identifying and Analyzing Prevalent AI Narratives with VET

Using the VET Framework, I show how different combinations of of valence, effectiveness, and trajectory identify common narratives about AI, and use the framework to compare and contrast these narratives. Attitudes about AI effectiveness and AI trajectory are correlated for many narratives (with Normalcy as a noteworthy exception that illustrates the value of considering perspectives on effectiveness and trajectory separately), so for simplicity Figure 1 depicts the space of AI attitudes as a two by two grid, with valence on the y-axis and a combined effectiveness/trajectory dimension on the x-axis. The four segments of this grid correspond to the four prevalent media attitudes in current U.S. discourse: AI Hype, AI Doom, AI Denial, and AI Normalcy.

### 3.1 AI Hype

*AI Hype* narratives combine a positive valence for AI with predictions of strong effectiveness and a rapid trajectory of progress, adoption, and impact. Media perspectives aligned with AI Hype imply that Artificial General Intelligence (AGI), a powerful form of AI that can meet or surpass human performance across nearly all tasks (Morris et al., 2025), will arrive within a short timeframe of a few

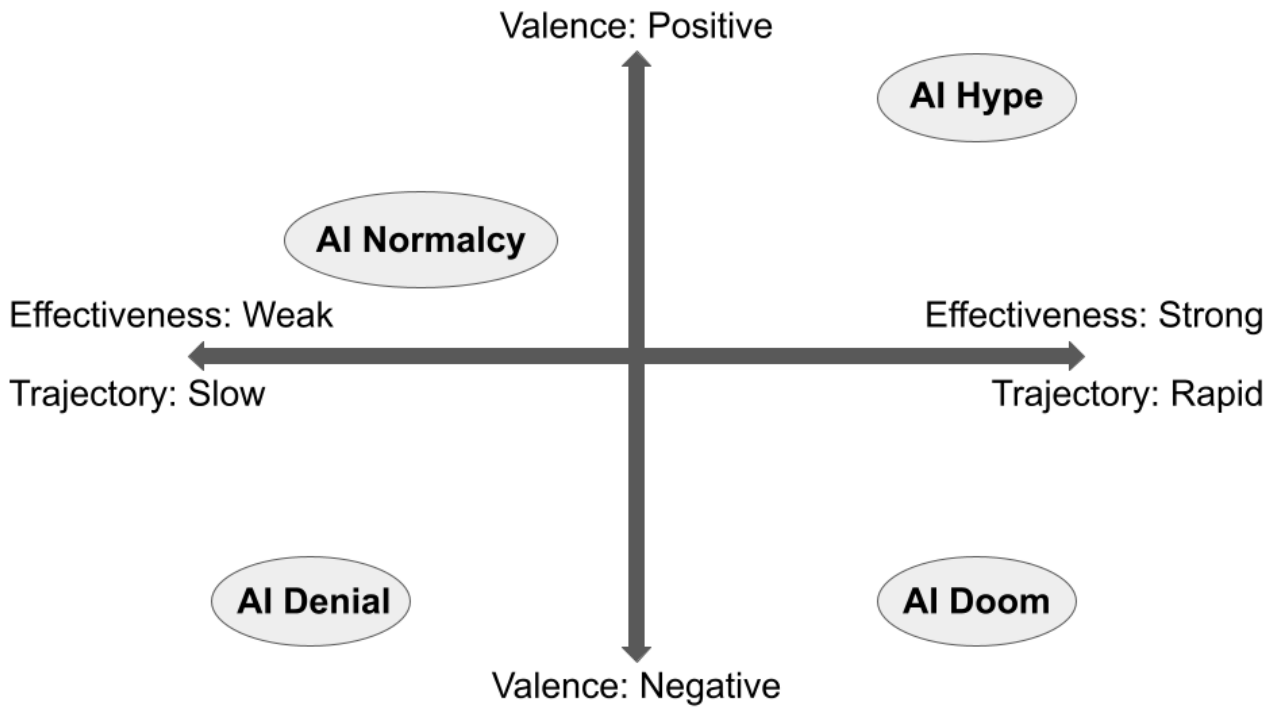


Figure 1. The VET Framework: An illustration of how the proposed analytic framework of valence, effectiveness, and trajectory helps to identify and compare four prevalent narratives in U.S. discourse about AI. Note that, per the framework, AI Normalcy is less strongly polarized with respect to valence and effectiveness than the other three narratives. Because effectiveness and trajectory are correlated in many narratives (with Normalcy as a counterexample), for simplicity the framework is visualized as a 2D grid.

years. AI Hype predicts that this powerful technology will be quickly deployed to usher in a utopian era of harmony and leisure, in which the power of AGI is used to address all of society’s ills, cure all disease, and create economic prosperity such that there is a superabundance of benefits for all the world’s citizens (Altman, 2024; Kurzweil, 2005).

AI Hype narratives are frequently associated with the leaders of AI startups and the venture capitalists who fund them, including OpenAI CEO Sam Altman’s essay “The Gentle Singularity” (Altman, 2025), Anthropic CEO Dario Amodei’s blog post “Machines of Loving Grace” (Amodei, 2024), venture capitalist Vinod Khosla’s essay “A Roadmap to AI Utopia” (Khosla, 2024), and venture capitalist Marc Andreessen’s blog post “The Techno-Optimist Manifesto” (Andreessen, 2023). Similarly, technology entrepreneur Elon Musk has made predictions that within about a decade, work and money will be made irrelevant by advances in AI and robotics (Rogelberg, 2026).

This techno-utopian narrative is reflected in the founding charters of several prominent frontier AI companies – OpenAI was founded as a non-profit with the goal of creating AGI so that it could distribute its benefits to all of humanity (OpenAI, 2018), Anthropic was founded as a public benefit corporation to ensure that transformative AI would positively benefit humanity (Anthropic), and DeepMind (now Google DeepMind) was founded with a goal of solving intelligence to benefit humanity (Mallaby, 2026).

Some variants of the AI Hype narrative assert not only that we are rapidly approaching AGI that will positively transform society, but also that these systems have become so advanced that they are, or may soon be, conscious. Philosopher David Chalmers (Long et al., 2024), former OpenAI Chief Scientist Ilya Sutskever (Sutskever, 2022), Google VP Blaise Agüera y Arcas (Agüera y Arcas, 2022), Nobel

Laureate Geoffery Hinton (Hinton, 2025b), and other prominent figures have all indicated they think that AI consciousness is possible or even likely. Anthropic has hired an ethicist to investigate the “AI welfare” of its Claude AI model (Anthropic, 2025b), taking actions such as cutting off angry user interactions with Claude in case Claude is an entity who may be emotionally harmed by users’ vitriolic inputs; Anthropic has also evolved Claude’s Constitution (Askell et al., 2026), a document of values and rules that guide Claude’s operation, to reflect the attitude of a parent conveying ethical norms to its child (Perrigo, 2026).

AI Hype narratives in the media can have a negative impact on the public’s AI Literacy by leading them to overtrust the capabilities of current AI systems and overestimate the rate at which technologies will be improved and adopted. Narratives that hype only positive impacts of AI with utopian visions of superabundance risk preventing the public from offering appropriate critiques of AI and developing systems and/or regulations to mitigate potential negative impacts.

Narratives that exaggerate the effectiveness of current systems as being akin to (or well beyond) human capabilities risk users’ misunderstanding important differences between AI systems and humans, and may result in overreliance (Morris et al., 2026), the formation of parasocial emotional relationships (Gabriel et al., 2024), or even the formation of diety-like beliefs about AI (Morris & Brubaker, 2025), among other risky behaviors. For example, today’s models exhibit jagged spikes and valleys in abilities, with a very different distribution of strengths and weaknesses than human populations (Morris et al., 2026); attributing human-like distributions of skills can mislead users to overgeneralize from strong performance in some areas, resulting in misapplying current AI tools to areas for which they are not yet well-suited. Further, narratives predicting extremely fast trajectories to superintelligence and thus superabundance may encourage risky behaviors such as changes in saving, investing, education, and/or employment strategies (Quiroz-Gutierrez, 2026).

Hype narratives assume an extremely fast trajectory of AI adoption. However, the trajectory of self-driving cars offers a cautionary counter-narrative. While robo-taxi services like Waymo are beginning to be deployed in select cities, largescale rollout of self-driving vehicles has been much slower than initially predicted even though the technology is available and has safety track records outperforming human drivers (Waymo, 2025). Negative public attitudes about self-driving vehicles, including concerns that they steal jobs from human drivers, misunderstandings about their safety records, and expectations that they should be held to a higher standard than human drivers, as well as regulatory and infrastructure challenges, have all slowed the deployment of this technology (American Automobile Association, 2025; Lee, 2025; West, 2024); a similar fate may befall modern general AI systems, thus slowing the adoption trajectory.

In the short to medium term, adoption trajectories are bottlenecked by limits on computation such as chip shortages (Sanders et al., 2026) and the need for data center and power grid infrastructure (International Energy Agency, 2025; Schneider, 2024). Such trajectories are likely even slower in areas such as the Global South, which often lack the infrastructure to run and utilize modern AI at scale (World Bank, 2025). National security concerns may also cause governing bodies to limit access to powerful future AI models (Hendrycks et al., 2025), thus limiting their ability to directly and uniformly benefit all of humanity. Some AI organizations and economists (Anthropic, 2025a; OpenAI, 2026b) have put forth proposals for redistributing the wealth AI is forecast to create via mechanisms such as Universal Basic Income, sovereign wealth funds, and/or robot taxes; however, there is currently no concrete public indication that the U.S. government has any serious near-term plans to implement such policy ideas, suggesting that the potential for a fast and positive financial impact trajectory for the general public is likely slower than Hype narratives predict.

Further, a key aspect of the impact trajectory assumed in Hype narratives is the assumption that technology can solve humanity’s problems; however, many problems do not have simple technological solutions, and instead are social or socio-technical in nature. For example, the invention of new mRNA vaccines for COVID-19 did not “solve” the COVID pandemic – some who wanted the vaccine could not get access to it (including many countries in the Global South), and some people who did have access chose not to take the vaccine. Technology alone does not necessarily solve all complex problems.

The attribution of extreme positive valence to AI in Hype discourse is called into question by the AI safety community. Indeed, the challenge of how to ensure that AI is aligned with human values remains an open area of inquiry (Christian, 2020). Additionally, for AI to be viewed positively at a global scale would require *pluralistic alignment* (Sorensen et al., 2024), *i.e.*, finding a way to align the values of AI that recognizes individual and cultural variance, since there is no universal definition of correct values. Successful pluralistic alignment remains aspirational as a goal; it is unclear that a single AI system could ever be viewed as a universal positive by diverse global communities with varied value systems.

Suggestions that current AI is conscious, present in some Hype narratives, are controversial and generally not widely accepted by scientists and philosophers (Comsa, 2026). Questions of whether future systems could have some form of consciousness are more ambiguous and a topic of much academic debate; however, current scientific methods cannot determine machine consciousness – indeed, the definition of consciousness in people in some situations such as newborns or those in a coma is still an open area of scientific inquiry (Seth, 2022). Focusing instead on how the *perception* of consciousness in AI systems comes about for end-users, and what impacts such attitudes may have, may be a more tractable area for public discussion, particularly as it relates to developing AI Literacy and supporting safe and responsible use of AI (Comsa, 2026).

### 3.2 AI Doom

*AI Doom* narratives emerge from assuming a negative valence for AI combined with strong effectiveness and rapid improvement trajectories. Portrayals of AI Doom in the media focus on existential and catastrophic risks of AI. Existential risks are those that would literally result in the extinction of the human race, and might result from a superintelligent AI system (Bostrom, 2014) that is unaligned with human values (Christian, 2020). Catastrophic risks focus on risks that might harm a large number of people, often through the misuse of powerful AI systems by human actors, such as to design and deploy advanced weapons (Hendrycks et al., 2025; Suleyman & Bhaskar, 2023; Kumar et al., 2025) or to commit serious cybercrimes (Carlini et al., 2026; Microsoft Threat Intelligence & OpenAI, 2024; Google Threat Intelligence Group, 2026).

A key proponent of the AI Doom narrative is Eliezar Yudkowsky, who founded the Machine Intelligence Research Institute (MIRI) and recently co-authored a book called *If Anyone Builds it, Everyone Dies* (Yudkowsky & Soares, 2025). Yudkowsky’s book conveys an extreme narrative of AI Doom, assigning a  $p(\text{doom})$  value of nearly 100%. Yudkowsky’s book uses a series of thought experiments to convince his readers that the creation of superintelligence would result in the extinction of humanity, essentially arguing that superintelligent computing systems would by definition develop their own goals, have the ability to defeat humans in order to advance those goals, and be so complex that humans would be unable to control them. This approach of using thought experiments to illustrate long-tail but extremely dangerous hypothetical scenarios around AI was popularized by the Oxford philosopher Nick Bostrom, whose influential book, *Superintelligence* (Bostrom, 2014), famously discusses the concept of a “paperclip maximizer” – a misaligned, superintelligent AI that destroys humanity in its quest to harvest all of the planet’s resources to create paperclips.

AI Doom perspectives have been espoused by more mainstream figures in the technology industry and academia, as well. One of the most influential proponents of Doom narratives is Geoffrey Hinton, a University of Toronto computer scientist and retired Google executive whose early research on neural network technology earned him the ACM Turing Award and the Nobel Prize in Physics; Hinton has given interviews in which he has indicated his own  $p(\text{doom})$  as 10%, 20%, or even 50% (Hinton, 2025a; Mallaby, 2026).

In 2023, many scientists and entrepreneurs signed a statement asserting that, “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war” (Center for AI Safety, 2023). Signatories included Microsoft co-founder Bill Gates, Google DeepMind CEO Demis Hassabis, Anthropic CEO Dario Amodei, OpenAI CEO Sam Altman, ACM Turing Award Winner Yoshua Bengio, and many other influential figures. Some signatories may not have intended to advocate fast-trajectory, high-probability AI Doom scenarios, but rather hoped to signal to policymakers that there was a non-zero risk of serious harms from AI, potentially over very long time horizons (Mallaby, 2026); however, media coverage amplified the negative valence and fast trajectory angle with headlines implying that credible authorities felt AI-caused human extinction may be near (e.g., (Roose, 2023)).

Media narratives that promulgate AI Doom may lead the public to overestimate the likelihood of extremely low-probability events and to overestimate the capabilities of AI. While considering a broad range of realistic and hypothetical outcomes of AI development is important for AI researchers and policymakers as part of a balanced portfolio of strategic planning and risk management, such narratives may confuse the general public about the near-term probabilities of existential and catastrophic risks.

The extreme negative valence of Doom narratives may result in AI backlash that causes people to overgeneralize about the risks of AI and reject safe and beneficial technologies, such as safe and robust applications of AI to medical, scientific, and cybersecurity advances. In extreme cases, Doom narratives might endanger the safety of those working on AI by inspiring members of the public to take violent action to mitigate hypothesized risks, as in the case of an individual who attacked the home of OpenAI CEO Sam Altman and attempted to set fire to the headquarters of OpenAI in early 2026 (Reuters, 2026).

By encouraging the public to focus on long-term, long-tail risk scenarios, Doom narratives might negatively impact AI Literacy by usurping focus from more pressing critiques of AI technologies that are likely to be more relevant to members of the general public, such as how AI may impact their physical and mental health, their interpersonal relationships, their education and labor opportunities, the environment, and geopolitics (Hanna & Bender, 2023) – however, current evidence suggests that while Doom narratives do increase the public’s fears about x-risk, they do not seem to lower people’s concern regarding other AI risks (Hoes & Gilardi, 2025).

Ironically, Doom narratives may also lead people to overtrust AI, because these narratives position AI as being extremely effective. Many Doom narratives also attribute consciousness to current or future AI systems, suggesting they may have their own goals that they choose to pursue at the expense of harming humankind; such narratives may encourage the public to attribute consciousness to AI systems, even when such determinations are not warranted by current scientific consensus (Comsa, 2026).

### 3.3 AI Denial

*AI Denial* narratives result from the combination of ascribing a negative valence to AI, low effectiveness, and slow improvement trajectories. AI Denial argues that AI technologies are not and can never be “intelligent” in the same way as humans, and therefore cannot work as promised by their developers.

Further, Denial narratives ascribe an extremely negative valence to AI, often arguing that the technology will increase societal inequalities, either by increasing income inequality (with the profits from AI going to technology companies or startup founders while others may lose their jobs to low-performing systems that do not work as well as the humans they replaced) or by increasing the gaps between favored and marginalized societal groups (*i.e.*, due to AI fairness concerns such as the use of AI to automate policing, access to social services, hiring processes, and other areas of life in which latent biases in such systems may disadvantage historically marginalized identity groups (Gallegos et al., 2024; Whittaker et al., 2018)). Note that these arguments appear to be in conflict with each other – many Denial narratives simultaneously suggest that AI is ineffective, yet nonetheless may be adopted too quickly, thereby causing harm. AI Denial arguments often suggest that the financial, environmental, and/or attentional resources society is directing toward AI development are being wasted on an undesirable and ineffective technology (Bender et al., 2021; Crawford, 2021; Narayanan & Kapoor, 2024; Thais, 2024).

An example of this narrative can be found in linguist Emily Bender’s and sociologist Alex Hanna’s book, *The AI Con* (Bender & Hanna, 2025), which uses semantic arguments to suggest that AI is a “con” because it cannot ever have the richness of human intellect and experience, arguing instead that modern AI systems are merely “stochastic parrots” (Bender et al., 2021) that use statistical processes to present an illusion of knowledge via fluent text production with no true understanding. Indeed, they argue against the use of the term “artificial intelligence” as being intentionally misleading to consumers, and instead they use alternate terminology such as “synthetic text-extruding machines” and even “racist pile of linear algebra” – this latter term is meant to reinforce the negative valence they ascribe to AI, *i.e.*, the idea that it is dangerous for society to adopt such systems, which may be ineffective such as by encoding racial biases learned from training data.

Another representative of the AI Denial viewpoint is Gary Marcus, an emeritus professor of psychology and neural technologies at NYU, whose substack *Marcus on AI* (Marcus, 2022) includes AI Denial commentary. Marcus’ posts have provocative titles such as, “Turns out Generative AI was a scam” (Marcus, 2026) and “Why the collapse of the Generative AI bubble may be imminent” (Marcus, 2024). Unlike Bender and Hanna, who take issue with the idea that any computational system can possess intelligence, Marcus’ form of Denial is more focused – he believes that the current technical approach underpinning modern generative AI systems will never succeed and therefore assumes weak effectiveness and a slow technical trajectory. Today’s systems are primarily driven by taking the transformer architecture (Vaswani et al., 2023) and applying scaling laws (Kaplan et al., 2020) with large amounts of data and compute, along with reinforcement learning (Ouyang et al., 2022; Silver et al., 2017), whereas Marcus argues that neuro-symbolic methods that explicitly encode knowledge and logical reasoning would be a more fruitful technical approach (Marcus & Davis, 2019).

AI Denial narratives threaten the public’s development of AI Literacy by encouraging the public to ignore this important class of emerging technologies because they argue that they have extremely low effectiveness and that the technical trajectory to improvement will also be quite slow. Indeed, some AI Denial commentary actively discourages the public from using technologies such as LLM-powered chatbots (Bender & Hanna, 2025), which may cause those who consume Denial narratives to lack knowledge and skills, thereby growing digital divides.

While current AI systems are imperfect, the Denial narrative of extremely low effectiveness ignores evidence that AI offers utility for many people and many application areas. ChatGPT was the most quickly adopted technology in history, with the market for LLM-powered chatbot users growing from zero to one hundred million active users within two months (Hu, 2023). AI-powered coding tools have changed the coding industry – established technology companies and startups alike now strongly

encourage or even mandate the use of agentic coding tools because of their proven utility, evidenced by recent systems scoring 87.6% on SWE-Bench (Anthropic, 2026), with senior leaders at several companies stating that high percentages of their employees' code is written with or by AI. Modern AI systems have led to scientific advances such as drug discovery (Bai et al., 2023; Mullin, 2026) and novel solutions to longstanding challenges in mathematics (Alexeev et al., 2026; OpenAI, 2026a; Alon et al., 2026). AI-powered software helps doctors spend more time with patients by performing automated note-taking (Tierney et al., 2025), and supports in analyzing medical data and imagery, allowing doctor-AI teams to be more effective at diagnosis than doctors working alone (Krakowski et al., 2024; Gommers et al., 2026). AI is able to identify serious cybersecurity issues in widely available software that human security experts could not find (Carlini et al., 2026; Microsoft Threat Intelligence & OpenAI, 2024; Google Threat Intelligence Group, 2026). These are just a few examples that illustrate the fallacy of the AI Denial narrative regarding effectiveness of current systems – while there are many things to improve about AI, to suggest that it does not offer any value is inaccurate.

AI Denial often suggests that individuals and society should not adopt AI because it does not have perfect accuracy or reliability. For example, AI systems might make mistakes, perhaps even mistakes that are systematically biased against certain demographic groups, due to biases in training data, such as in medical diagnosis or resource allocation (Whittaker et al., 2018). However, choosing appropriate baselines for comparison is important to analyzing effectiveness. The threshold for AI deployment for most use cases need not be perfection, but comparison to *current, realistic baselines* for correctness and cost. Human doctors make mistakes in diagnosis; if an AI diagnosis system makes mistakes but makes fewer mistakes than human doctors, this might be of great benefit to many people, even if it is not perfect. Human policymakers exhibit bias (conscious and unconscious) among various demographic groups in resource allocation; even AI systems that have some residual bias may in practice have less biased outcomes than current processes, and such biases may be easier to audit than latent biases in human decision makers (Kleinberg et al., 2018). Alternatively, society may be willing to accept small tradeoffs in accuracy for some tasks if costs are lowered (Li et al., 2026). Helping the public understand the baselines for comparison for judging AI's effectiveness, and the dimensions to consider when making decisions about whether to deploy AI (accuracy, costs, baselines, tradeoffs) is a critical component of AI Literacy that is not conveyed through Denial narratives that demand perfection before adoption.

The focus of some AI Denial narratives on the semantics of the term “Artificial Intelligence” is perhaps well-intentioned from an AI Literacy standpoint, but is largely moot from a practical standpoint. Artificial Intelligence has become a “term of art,” and most academics and companies in the AI space do not make any claims that AI is literally “intelligent” in the same way as a human; rather, most focus on the practical capabilities of AI systems rather than amorphous definitions of the semantics of the word “intelligence.” However, conveying that AI, while useful for many tasks, need not be conflated with human intelligence or with human-like depictions of AI from science fiction should be an important component of responsible AI discourse.

### 3.4 AI Normalcy

*AI Normalcy* narratives have neutral-to-positive assumptions about AI valence, but combine this with moderate effectiveness and slow improvement, adoption, and impact trajectories. Essentially, AI Normalcy suggests that today's improvements in AI should not be viewed differently than past innovations, that they will not radically disrupt society, and that therefore there is no urgency for individuals, communities, organizations, or governments to take drastic actions regarding AI preparedness.

A representative example of this narrative is Princeton computer scientists' Arvind Narayanan's and Sayash Kapoor's 2025 essay, "AI as Normal Technology" (Narayanan & Kapoor, 2025), which argues that while modern AI technologies have made some noteworthy progress, these technologies should be viewed similarly to other past innovations such as electricity or personal computers, which required decades to reach mass adoption and demonstrate concrete productivity gains. MIT Robotician Rodney Brooks is another prominent AI Normalcy advocate, arguing that short term predictions about technologies' impacts tend to be overblown, whereas long-term adoption and uses (over the course of decades) can be unanticipated (Brooks, 2017).

Economic analyses that anticipate labor replacement due to AI as likely to be relatively slow (Brynjolfsson et al., 2021), that any job losses will be offset by gains from new classes of AI-enabled jobs as has been the pattern with past technologies (Autor, 2015; World Economic Forum, 2025), or that AI-enhanced productivity will drive further demand for labor rather than replace it (i.e., a Jevons Paradox) (Jevons, 1865) are also AI Normalcy positions. Invocation of the Solow Paradox (the idea that "You can see the computer age everywhere but in the productivity statistics") (Triplett, 1999) is another example of an AI Normalcy stance.

The AI Normalcy narrative thus argues that adoption and impact trajectories are likely to be slow, due to the need for individuals and businesses to understand how to integrate novel AI technologies into their workflows effectively, for the technologies to become reliable and cheap enough that it is cost effective to use them at scale, and due to the likely need to overcome societal frictions to fast adoption such as those introduced by labor unions, regulators, or public backlash.

Note that AI Normalcy is less "extreme" per the VET framework than the Hype, Doom, and Denial narratives. In particular, the Normalcy narrative tends to view effectiveness as moderate rather than weak, and is relatively neutral regarding the valence of AI. The Normalcy narrative assumes a very slow adoption and impact trajectory, however, as well as a relatively slow technical trajectory – for instance, expressing skepticism that AGI or superintelligence are meaningfully defined terms and/or realistic technical goals (Narayanan & Kapoor, 2025).

This nuanced analysis of Normalcy narratives demonstrates how the VET Framework goes beyond prior taxonomies of AI perspectives such as Gilardi et al.'s ontology of existential risk, accelerationist, immediate risk, and balanced risk (Gilardi et al., 2024) – AI Normalcy arguments would likely be categorized as "balanced risk" in their framework, suggesting such views are the desired and accurate narrative. However, with the VET framework, we see that while AI Normalcy is less extreme in its positioning regarding valence and effectiveness than other prevalent narratives, it nonetheless poses risks to AI Literacy because of its slow positioning along the trajectory dimension. Normalcy narratives risk a public that is unprepared for the rate and degree of personal and societal changes that many experts expect AI to bring (Murphy et al., 2025).

People are generally bad at understanding exponential growth curves (Wagenaar & Sagaria, 1975); the Normalcy narrative projects a non-exponential adoption and impact trajectory of AI, in line with previous technologies. However, current evidence suggests this view of AI's trajectory is inaccurate. For example, the recent case of Anthropic's Mythos model needing to be withheld because it was able to identify a large number of critical cybersecurity flaws (Carlini et al., 2026) is an example of AI creating a serious and rapid step-change in cyber capabilities, requiring national security responses and potentially having serious geopolitical implications. The recent successes of AI in discovering novel solutions to longstanding open challenges in mathematics (Alexeev et al., 2026; OpenAI, 2026a; Alon et al., 2026) is another hint that this technology is on a much steeper impact trajectory than Normalcy

suggests.

Indeed, another piece of evidence that Normalcy discourse likely underestimates the technical trajectory of AI is that frontier AI labs are openly pursuing a strategy of *recursive self-improvement* (Zelikman et al., 2024) for AI research and development, in which they invest in optimizing AI tools for writing code and performing AI research. The goal of recursive self-improvement is to bootstrap the rate of AI development, potentially enabling a *fast takeoff* (Bostrom, 2014) or *intelligence explosion* (Good, 1965) of AI capability since AI that can speed up its own development will logically accelerate future AI advances. This strategy is already bearing fruit, as evidenced by the acceleration of utility of agentic coding tools (Carlini et al., 2026; Anthropic, 2026), which have now been widely adopted throughout the AI industry. The impacts of these changes are being reflected not only in the pace of AI development, but in changes to the labor market for software developers, including substantial layoffs at some software companies that are choosing to rely more on AI coding tools (Angelo, 2026; Ruwitch, 2026) and a slowdown in hiring for junior coding positions (Brynjolfsson et al., 2025a). Such developments provide counter-evidence to the narrative that AI will mature and be adopted at a rate on par with prior technologies, providing early signals of a faster-than-anticipated adoption trajectory in some sectors.

While bottlenecks to adoption certainly exist, including social issues such as AI backlash among some segments of the general public (McClain et al., 2025), potential future changes in the regulatory landscape, and Amdahl's Law (Amdahl, 1967) bottlenecks on the extent to which AI alone can speed up particular industries or processes, the Normalcy narrative risks fomenting complacency among the public and policymakers by leading them to underestimate the time they may have available to prepare for and adapt to likely impacts of AI. This might include both personal preparedness decisions such as re-skilling and collective preparedness decisions such as advocating for particular policies or regulations.

The VET framework also offers insight into why Normalcy narratives may be less prevalent in the media ecosystem than Hype, Doom, or Denial. Traditional and social media tend to amplify more extreme narratives and narratives with negative valence (Rathje et al., 2021; Brady et al., 2017; Robertson et al., 2023) – because Normalcy is less extreme in its predictions about the valence, effectiveness, and trajectory of AI than competing narratives, it may receive less media traction.

## 4 Conclusion

In this article, I introduced the VET Framework, a new tool for supporting AI Literacy by providing a method of analyzing AI discourse using the dimensions of valence, effectiveness, and trajectory. *Valence* reflects whether an AI narrative depicts AI as negative or positive, including considerations of AI's current and future impacts, meta-issues such as attitudes toward the technology industry, privacy, and capitalism, and even the positive or negative motivations that a hypothesized sentient AI might itself possess. *Effectiveness* encompasses whether an AI narrative positions AI as weak or strong, including considerations of the accuracy and reliability of AI systems, potentially with respect to human baselines. *Trajectory* considers whether a narrative positions AI progress as slow or rapid, including the technical trajectory for AI progress, the adoption trajectory of AI by individuals and organizations, and the rate at which AI will impact society.

In this article, I have demonstrated how the framework allows for identifying, comparing, and analyzing prevailing media narratives about AI by using the framework to characterize AI Hype, AI Doom, AI Denial, and AI Normalcy based on their depictions of AI's valence, effectiveness, and trajectory.

This framework can support AI Literacy by offering a method to systematically “vet” the quality of arguments about AI using these three analytical dimensions. I hope that the VET Framework can be a useful tool for policymakers, the media, and the general public to support nuanced discussion and analysis of both the current state of AI and its likely future.

## Acknowledgements

This article benefited from a fellowship at the Paris Institute for Advanced Study (France), with the financial support of the French State, programme *Investissements d’avenir* managed by the Agence Nationale de la Recherche (ANR-11-LABX-0027-01 Labex RFIEA+).

I would like to thank John Krumm, Siddharth Suri, Iason Gabriel, Michael Terry, Jacy Reese Anthis, Shriya Sekhsaria, Pierre Dillenbourg, and Jan Eißfeldt for helpful conversations and feedback that influenced this work. I would especially like to thank Dan Morris, whose support made my writing residency at Paris IAS possible.

Note that this work represents the opinion of the author, and is not an official policy statement or position of any institutions with which she is associated.

## References

- Acemoglu, D. The simple macroeconomics of AI. *Economic Policy*, 40(121):13–58, 2025. doi: 10.1093/epolic/eiae042.
- Agüera y Arcas, B. Do large language models understand us? *Daedalus*, 151(2):183–197, 2022. doi: 10.1162/daed\_a\_01909.
- Alexeev, B., Barreto, K., Li, Y., Lichtman, J. D., Price, L., Shah, J. I., Tang, Q., and Tao, T. Primitive sets and von Mangoldt chains: Erdős problem #1196 and beyond, 2026. URL <https://arxiv.org/abs/2605.00301>.
- Alon, N., Bloom, T. F., Gowers, W. T., Litt, D., Sawin, W., Shankar, A., Tsimmerman, J., Wang, V., and Wood, M. M. Remarks on the disproof of the unit distance conjecture, 2026. URL <https://arxiv.org/abs/2605.20695>.
- Altman, S. The intelligence age. <https://ia.samaltman.com>, September 2024.
- Altman, S. The gentle singularity, June 2025. URL <https://blog.samaltman.com/the-gentle-singularity>.
- Amdahl, G. M. Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the AFIPS Spring Joint Computer Conference*, AFIPS ’67 (Spring), pp. 483–485, Reston, VA, 1967. AFIPS Press.
- American Automobile Association. Fear in self-driving vehicles persists. AAA Newsroom. <https://newsroom.aaa.com/2025/02/aaa-fear-in-self-driving-vehicles-persists/>, February 2025.
- Amodei, D. Machines of loving grace, October 2024. URL <https://darioamodei.com/essay/machines-of-loving-grace>.
- Andreessen, M. The techno-optimist manifesto. Andreessen Horowitz. <https://a16z.com/the-techno-optimist-manifesto/>, October 2023.

- Angelo, J. Block CEO Jack Dorsey lays off nearly half of his staff because of AI and predicts most companies will make similar cuts in the next year. *Fortune*. <https://fortune.com/2026/02/27/block-jack-dorsey-ceo-xyz-stock-square-4000-ai-layoffs/>, February 2026.
- Anthropic. Company. <https://www.anthropic.com/company>. Accessed May 21, 2026.
- Anthropic. Preparing for AI's economic impact: Exploring policy responses. <https://www.anthropic.com/research/economic-policy-responses>, October 2025a.
- Anthropic. Exploring model welfare. <https://www.anthropic.com/research/exploring-model-welfare>, April 2025b.
- Anthropic. Introducing Claude Opus 4.7. <https://www.anthropic.com/news/claude-opus-4-7>, April 2026.
- Askell, A., Carlsmith, J., Olah, C., Kaplan, J., Karnofsky, H., and many other contributors. Claude's constitution. Anthropic. <https://www.anthropic.com/constitution>, January 2026.
- Autor, D. H. Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30, August 2015. doi: 10.1257/jep.29.3.3.
- Axios and The Harris Poll. 2026 Axios Harris Poll 100: GOP embraces AI over democrats. Axios, May 2026. URL <https://www.axios.com/2026/05/19/axios-harris-poll-100-ai-politics>.
- Bai, F., Li, S., and Li, H. AI enhances drug discovery and development. *National Science Review*, 11(3):nwad303, 2023. doi: 10.1093/nsr/nwad303.
- Bender, E. M. and Hanna, A. *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. Harper, May 2025.
- Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pp. 610–623, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445922. URL <https://doi.org/10.1145/3442188.3445922>.
- Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., and Bavel, J. J. V. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28): 7313–7318, 2017. doi: 10.1073/pnas.1618923114. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1618923114>.
- Briggs, J. and Kodnani, D. The potentially large effects of artificial intelligence on economic growth. Technical report, Goldman Sachs Economics Research, March 2023. URL <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.
- Brooks, R. The seven deadly sins of AI predictions, October 2017. URL <https://www.technologyreview.com/2017/10/06/241837/the-seven-deadly-sins-of-ai-predictions/>.

- Brynjolfsson, E., Rock, D., and Syverson, C. The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372, 2021. doi: 10.1257/mac.20180386.
- Brynjolfsson, E., Chandar, B., and Chen, R. Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence. Technical report, Stanford Digital Economy Lab, November 2025a. URL <https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/>.
- Brynjolfsson, E., Li, D., and Raymond, L. Generative AI at work. *The Quarterly Journal of Economics*, 140(2):889–942, 2025b. doi: 10.1093/qje/qjae044.
- Carlini, N., Cheng, N., Lucas, K., Moore, M., Nasr, M., Prabhushankar, V., Xiao, W., Angulu, H., Ben Asher, E., Bow, J., Bradwell, K., Buchanan, B., Forsythe, D., Freeman, D., Gaynor, A., Ge, X., Graham, L., Guru, K., Lakhani, H., McNiece, M., Mehrara, M., Nichol, R., Pirzada, A., Porter, S., Terzis, A., and Troy, K. Assessing Claude Mythos Preview’s cybersecurity capabilities. Anthropic Red Team Blog. <https://red.anthropic.com/2026/mythos-preview/>, April 2026.
- Center for AI Safety. Statement on ai risk, May 2023. URL <https://aistatement.com/#open-letter>.
- Christian, B. *The Alignment Problem: Machine Learning and Human Values*. W.W. Norton & Company, 2020.
- Comsa, I.-M. Ai and consciousness: Shifting focus towards tractable questions, 2026. URL <https://arxiv.org/abs/2605.06965>.
- Crawford, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, New Haven, CT, 2021. ISBN 9780300209570.
- De Freitas, J., Uğuralp, A. K., Uğuralp, Z., and Puntoni, S. AI companions reduce loneliness. *Journal of Consumer Research*, 2025. doi: 10.1093/jcr/ucaf040.
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., and Agarwal, S. How ai and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study, 2025. URL <https://arxiv.org/abs/2503.17473>.
- Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., Mateos-Garcia, J., Weidinger, L., Street, W., Lange, B., Ingerman, A., Lentz, A., Enger, R., Barakat, A., Krakovna, V., Siy, J. O., Kurth-Nelson, Z., McCroskery, A., Bolina, V., Law, H., Shanahan, M., Alberts, L., Balle, B., de Haas, S., Ibitoye, Y., Dafoe, A., Goldberg, B., Krier, S., Reese, A., Witherspoon, S., Hawkins, W., Rauh, M., Wallace, D., Franklin, M., Goldstein, J. A., Lehman, J., Klenk, M., Vallor, S., Biles, C., Morris, M. R., King, H., y Arcas, B. A., Isaac, W., and Manyika, J. The ethics of advanced ai assistants, 2024. URL <https://arxiv.org/abs/2404.16244>.
- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Deroncourt, F., Yu, T., Zhang, R., and Ahmed, N. K. Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3):1097–1179, 2024. doi: 10.1162/coli\_a.00524.

- Gerlich, M. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1):6, 2025. doi: 10.3390/soc15010006.
- Gilardi, F., Kasirzadeh, A., Bernstein, A., Staab, S., and Gohdes, A. We need to understand the effect of narratives about generative AI. *Nature Human Behaviour*, 8(12):2251–2252, December 2024. ISSN 2397-3374. doi: 10.1038/s41562-024-02026-z. Published online 21 October 2024.
- Gommers, J., Hernström, V., Josefsson, V., Sartor, H., Schmidt, D., Hjelmgren, A., Larsson, A.-M., Hofvind, S., Andersson, I., Rosso, A., Hagberg, O., and Lång, K. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study: A randomised, controlled, non-inferiority, single-blinded, population-based, screening-accuracy trial. *The Lancet*, 407(10527):505–514, January 2026. doi: 10.1016/S0140-6736(25)02464-X.
- Good, I. J. Speculations concerning the first ultraintelligent machine. In Alt, F. L. and Rubinoff, M. (eds.), *Advances in Computers*, volume 6, pp. 31–88. Academic Press, New York, 1965. doi: 10.1016/S0065-2458(08)60418-0.
- Goodman, S. M., Buehler, E., Clary, P., Coenen, A., Donsbach, A., Horne, T. N., Lahav, M., MacDonald, R., Michaels, R. B., Narayanan, A., Pushkarna, M., Riley, J., Santana, A., Shi, L., Sweeney, R., Weaver, P., Yuan, A., and Morris, M. R. Lampost: Ai writing assistance for adults with dyslexia using large language models. *Commun. ACM*, 67(9):80–89, August 2024. ISSN 0001-0782. doi: 10.1145/3626952. URL <https://doi.org/10.1145/3626952>.
- Google Threat Intelligence Group. Adversaries leverage AI for vulnerability exploitation, augmented operations, and initial access. Google Cloud Blog. <https://cloud.google.com/blog/topics/threat-intelligence/ai-vulnerability-exploitation-initial-access>, May 2026.
- Gramlich, J., Kennedy, B., McClain, C., and Stocking, G. How Americans view data centers’ impact in key areas, from the environment to jobs. Technical report, Pew Research Center, March 2026. URL <https://www.pewresearch.org/short-reads/2026/03/12/how-americans-view-data-centers-impact-in-key-areas-from-the-environment-to-jobs/>.
- Gunder, A. Ai literacies in focus: From frameworks to action, 2025. URL <https://wcet.wiche.edu/wp-content/uploads/sites/11/2025/09/2025-WCET-AI-Literacies-Lit-in-Focus-1.pdf>.
- Hanna, A. and Bender, E. M. AI causes real harm. let’s focus on that over the end-of-humanity hype. *Scientific American*, 328:69, 2023.
- Hendrycks, D., Schmidt, E., and Wang, A. Superintelligence strategy: Expert version, 2025. URL <https://arxiv.org/abs/2503.05628>.
- Hinton, G. “Godfather of AI” Geoffrey Hinton warns AI could take control from humans. CBS News. <https://www.cbsnews.com/news/godfather-of-ai-geoffrey-hinton-ai-warning>, 2025a.
- Hinton, G. Interview on *Tonight with Andrew Marr*. LBC. [https://www.lbc.co.uk/article/ai-consciousness-geoffrey-hinton-5HjdRXD\\_2/](https://www.lbc.co.uk/article/ai-consciousness-geoffrey-hinton-5HjdRXD_2/), January 2025b.

- Hoes, E. and Gilardi, F. Existential risk narratives about AI do not distract from its immediate harms. *Proceedings of the National Academy of Sciences*, 122(16):e2419055122, April 2025. doi: 10.1073/pnas.2419055122.
- Hu, K. Chatgpt sets record for fastest-growing user base - analyst note. *Reuters*, February 2023. URL <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.
- International Energy Agency. Electricity 2025: Analysis and forecast to 2027. Technical report, IEA, Paris, 2025. URL <https://www.iea.org/reports/electricity-2025>.
- Jevons, W. S. *The Coal Question: An Inquiry Concerning the Progress of the Britain and the Probable Exhaustion of Our Coal-Mines*. Macmillan, London, 1865.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling laws for neural language models, 2020. URL <https://arxiv.org/abs/2001.08361>.
- Khosla, V. A roadmap to AI utopia. *Time*. <https://time.com/7174892/a-roadmap-to-a-i-utopia/>, November 2024.
- Kleinberg, J., Ludwig, J., Mullainathan, S., and Sunstein, C. R. Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10:113–174, 12 2018. ISSN 2161-7201. doi: 10.1093/jla/laz001. URL <https://doi.org/10.1093/jla/laz001>.
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., and Maes, P. Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task, 2025. URL <https://arxiv.org/abs/2506.08872>.
- Krakowski, I., Kim, J., Cai, Z. R., Daneshjou, R., Lapins, J., Eriksson, H., Lykou, A., and Linos, E. Human-AI interaction in skin cancer diagnosis: A systematic review and meta-analysis. *npj Digital Medicine*, 7:78, April 2024. doi: 10.1038/s41746-024-01031-w.
- Kumar, D., Birur, N. A., Baswa, T., Agarwal, S., and Harshangi, P. Quantifying cbrn risk in frontier models, 2025. URL <https://arxiv.org/abs/2510.21133>.
- Kurzweil, R. *The Singularity Is Near: When Humans Transcend Biology*. Viking, New York, 2005. ISBN 9780670033843.
- Laupichler, M. C., Aster, A., Schirch, J., and Raupach, T. Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3: 100101, 2022. ISSN 2666-920X. doi: <https://doi.org/10.1016/j.caeai.2022.100101>. URL <https://www.sciencedirect.com/science/article/pii/S2666920X2200056X>.

- Lee, T. B. Unions want to ban driverless taxis—will democratic leaders say yes? Understanding AI. <https://www.understandingai.org/p/unions-want-to-ban-driverless-taxiswill>, August 2025.
- Lenarczyk, G., Minssen, T., Price, N., and Rai, A. The future of ai regulation in drug development: a comparative analysis. *Journal of Law and the Biosciences*, 12(2):lsaf028, 07 2025. ISSN 2053-9711. doi: 10.1093/jlb/lsaf028. URL <https://doi.org/10.1093/jlb/lsaf028>.
- Li, P., Yang, J., Islam, M. A., and Ren, S. Making AI less “thirsty”: Uncovering and addressing the secret water footprint of AI models. *Communications of the ACM*, 2025. doi: 10.1145/3724499.
- Li, W., Aboutorabi, A., Lyu, H., Qian, K., Fleming, M., Goehring, B. C., and Thompson, N. Economics of human and ai collaboration: When is partial automation more attractive than full automation?, 2026. URL <https://arxiv.org/abs/2603.29121>.
- Long, D. and Magerko, B. What is ai literacy? competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, pp. 1–16, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi: 10.1145/3313831.3376727. URL <https://doi.org/10.1145/3313831.3376727>.
- Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., Pfau, J., Sims, T., Birch, J., and Chalmers, D. Taking ai welfare seriously, 2024. URL <https://arxiv.org/abs/2411.00986>.
- Lu, C., Lu, C., Lange, R. T., Yamada, Y., Hu, S., Foerster, J., Ha, D., and Clune, J. Towards end-to-end automation of AI research. *Nature*, 651(8107):914–919, 2026. doi: 10.1038/s41586-026-10265-5.
- Mallaby, S. *The Infinity Machine: Demis Hassabis, DeepMind, and the Quest for Superintelligence*. Penguin Press, New York, 2026. ISBN 9780593831847.
- Marcus, G. Marcus on AI. Substack newsletter. <https://garymarcus.substack.com/>, 2022.
- Marcus, G. Why the collapse of the generative AI bubble may be imminent. Marcus on AI (Substack), August 2024. URL <https://garymarcus.substack.com/p/why-the-collapse-of-the-generative-ai-bubble-may-be-imminent>.
- Marcus, G. Turns out generative AI was a scam. Marcus on AI (Substack), February 2026. URL <https://garymarcus.substack.com/p/turns-out-generative-ai-was-a-scam>.
- Marcus, G. and Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon, September 2019.
- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Echemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Betts Lotufo, J., Rome, A., Shi, A., and Oak, S. Artificial intelligence index report 2025. Technical report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, 2025. URL <https://hai.stanford.edu/ai-index/2025-ai-index-report>.
- McClain, C., Kennedy, B., Gottfried, J., Anderson, M., and Pasquini, G. How the US public and AI experts view artificial intelligence. Technical report, Pew Research Center, April 2025. URL <https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/>.

- Miao, F., Shiohira, K., and Lao, N. Ai competency framework for students, August 2024. URL <https://www.unesco.org/en/articles/ai-competency-framework-students>.
- Microsoft Threat Intelligence and OpenAI. Staying ahead of threat actors in the age of AI. Technical report, Microsoft, February 2024. URL <https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/>.
- Morris, M. R. and Brubaker, J. R. Generative ghosts: Anticipating benefits and risks of ai afterlives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25*, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713758. URL <https://doi.org/10.1145/3706598.3713758>.
- Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., and Legg, S. Levels of agi for operationalizing progress on the path to agi, 2025. URL <https://arxiv.org/abs/2311.02462>.
- Morris, M. R., Altman, D., Belfield, H., Goemans, A., Iqbal, H., Burnell, R., Gabriel, I., Albanie, S., and Dafoe, A. Characterizing model jaggedness supports safety and usability. 2026. URL [https://cs.stanford.edu/~merrie/papers/jaggedness\\_preprint.pdf](https://cs.stanford.edu/~merrie/papers/jaggedness_preprint.pdf).
- Mullin, E. AI-designed drugs by a DeepMind spinoff are headed to human trials. *Wired*. <https://www.wired.com/story/wired-health-2026-how-ai-is-powering-drug-discovery-max-jaderberg/>, April 2026.
- Murphy, C., Rosenberg, J., Canedy, J., Jacobs, Z., Flechner, N., Britt, R., Pan, A., Rogers-Smith, C., Mayland, D., Buffington, C., Kučinskis, S., Coston, A., Kerner, H., Pierson, E., Rabbany, R., Salganik, M., Seamans, R., Su, Y., Tramèr, F., Hashimoto, T., Narayanan, A., Tetlock, P. E., and Karger, E. The longitudinal expert ai panel: Understanding expert views on ai capabilities, adoption, and impact. Working paper 5, Forecasting Research Institute, 2025. URL <https://leap.forecastingresearch.org/reports/wave3>.
- Mügge, D. Eu ai sovereignty: for whom, to what end, and to whose benefit? *Journal of European Public Policy*, 31(8):2200–2225, 2024. doi: 10.1080/13501763.2024.2318475.
- Narayanan, A. and Kapoor, S. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton University Press, Princeton, NJ, 2024. ISBN 9780691249131.
- Narayanan, A. and Kapoor, S. Ai as normal technology, April 2025. URL <https://knightcolumbia.org/content/ai-as-normal-technology>.
- Noy, S. and Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023. doi: 10.1126/science.adh2586.
- OECD. Bridging the ai skills gap: Is training keeping up?, 2025.
- OpenAI. OpenAI charter. <https://openai.com/charter/>, April 2018.
- OpenAI. Planar point sets with many unit distances, May 2026a. URL <https://cdn.openai.com/pdf/74c24085-19b0-4534-9c90-465b8e29ad73/unit-distance-proof.pdf>.
- OpenAI. Industrial policy for the intelligence age: Ideas to keep people first. <https://openai.com/index/industrial-policy-for-the-intelligence-age/>, April 2026b.

- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. Training language models to follow instructions with human feedback, 2022. URL <https://arxiv.org/abs/2203.02155>.
- Paglialunga, A. and Melogno, S. The effectiveness of artificial intelligence-based interventions for students with learning disabilities: A systematic review. *Brain Sciences*, 15(8):806, 2025. doi: 10.3390/brainsci15080806.
- Perrigo, B. Can you teach an AI to be good? Anthropic thinks so. *Time*. <https://time.com/7354738/claude-constitution-ai-alignment/>, January 2026.
- Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., and Maes, P. Investigating affective use and emotional well-being on chatgpt, 2025. URL <https://arxiv.org/abs/2504.03888>.
- Quiroz-Gutierrez, M. Elon musk says saving for retirement is irrelevant because AI is going to create a world of abundance: ‘it won’t matter’. *Fortune*. <https://fortune.com/2026/01/12/elon-musk-retirement-savings-irrelevant-ai-robots-abundance/>, January 2026.
- Rabanser, S., Kapoor, S., Kirgis, P., Liu, K., Utpala, S., and Narayanan, A. Towards a science of ai agent reliability, 2026. URL <https://arxiv.org/abs/2602.16666>.
- Rathje, S., Bavel, J. J. V., and van der Linden, S. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26):e2024292118, 2021. doi: 10.1073/pnas.2024292118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2024292118>.
- Reuters. Man charged after molotov cocktail attack on openai ceo sam altman’s home. *Reuters*, April 2026. URL <https://www.reuters.com/legal/government/man-charged-after-molotov-cocktail-attack-openai-ceo-sam-altmans-home-2026-04-13/>.
- Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., and Feuerriegel, S. Negativity drives online news consumption. *Nature Human Behaviour*, 7(5):812–822, May 2023. ISSN 2397-3374. doi: 10.1038/s41562-023-01538-4. URL <https://doi.org/10.1038/s41562-023-01538-4>.
- Rogelberg, S. Elon Musk says that in 10 to 20 years, work will be optional and money will be irrelevant thanks to AI and robotics. *Fortune*. <https://fortune.com/2026/01/19/when-does-elon-musk-say-work-will-be-optional-and-money-will-be-irrelevant-ai-robotics/>, January 2026.
- Roose, K. A.i. poses ‘risk of extinction,’ industry leaders warn. *The New York Times*. <https://www.nytimes.com/2023/05/30/technology/ai-chatbot-extinction-risk.html>, May 2023.
- Ruwitch, J. Meta slashes 8,000 jobs as it pivots towards AI. *NPR*. <https://www.npr.org/2026/05/20/nx-s1-5826917/meta-layoffs-ai-jobs>, May 2026.
- Sanders, J., Egan, J., and Madigan, R. American AI companies can’t get enough chips. Technical report, Center for a New American Security (CNAS), May 2026. URL <https://www.cnas.org/publications/reports/american-ai-companies-cant-get-enough-chips>.

- Schneider, J. AI to drive 165% increase in data center power demand by 2030. Goldman Sachs Research. <https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030>, 2024.
- Seth, A. *Being You: A New Science of Consciousness*. Faber & Faber, 2022.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, Oct 2017. ISSN 1476-4687. doi: 10.1038/nature24270. URL <https://doi.org/10.1038/nature24270>.
- Sorensen, T., Moore, J., Fisher, J., Gordon, M., Mireshghallah, N., Rytting, C. M., Ye, A., Jiang, L., Lu, X., Dziri, N., Althoff, T., and Choi, Y. A roadmap to pluralistic alignment, 2024. URL <https://arxiv.org/abs/2402.05070>.
- Suleyman, M. and Bhaskar, M. *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma*. Crown, New York, 2023. ISBN 9780593593950.
- Sutskever, I. “it may be that today’s large neural networks are slightly conscious”. Post on X (formerly Twitter). <https://x.com/ilyasut/status/1491554478243258368>, February 2022.
- Tadimalla, S. Y. and Maher, M. L. Ai literacy as a core component of ai education. *AI Magazine*, 46, July 2025. doi: <https://doi.org/10.1002/aaai.70007>.
- Thais, S. Misrepresented technological solutions in imagined futures: The origins and dangers of ai hype in the research community, 2024. URL <https://arxiv.org/abs/2408.15244>.
- Tierney, A. A., Gayre, G., Hoberman, B., Mattern, B., Balleca, M., Wilson Hannay, S., Castilla, K., Lau, C., Kipnis, P., Liu, V., and Lee, K. Ambient artificial intelligence scribes: Learnings after 1 year and over 2.5 million uses. *NEJM Catalyst*, 2025. doi: 10.1056/CAT.25.0040.
- Triplett, J. E. The solow productivity paradox: What do computers do to productivity? *Canadian Journal of Economics*, 32(2):309–334, April 1999. doi: 10.2307/136425.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need, 2023. URL <https://arxiv.org/abs/1706.03762>.
- Wagenaar, W. A. and Sagaria, S. D. Misperception of exponential growth. *Perception & Psychophysics*, 18(6):416–422, 1975. doi: 10.3758/BF03204114.
- Waymo. Safety impact hub. <https://waymo.com/safety/impact/>, 2025.
- West, D. M. The evolving safety and policy challenges of self-driving cars. Brookings Institution. <https://www.brookings.edu/articles/the-evolving-safety-and-policy-challenges-of-self-driving-cars/>, July 2024.
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Myers West, S., Richardson, R., Schultz, J., and Schwartz, O. AI now 2018 report. Technical report, AI Now Institute, New York, December 2018. URL [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.html](https://ainowinstitute.org/AI_Now_2018_Report.html).
- World Bank. Digital progress and trends report 2025. Technical report, World Bank, November 2025. URL <https://openknowledge.worldbank.org/entities/publication/8f5d2cb9-92d4-42fd-a4ad-fa538f081488>.

World Economic Forum. Future of jobs report 2025. Technical report, World Economic Forum, January 2025. URL [https://reports.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_Report\\_2025.pdf](https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf).

Yudkowsky, E. and Soares, N. *If Anyone Builds It, Everyone Dies: Why Superhuman AI Would Kill Us All*. Little, Brown and Company, September 2025.

Zelikman, E., Lorch, E., Mackey, L., and Kalai, A. T. Self-taught optimizer (stop): Recursively self-improving code generation, 2024. URL <https://arxiv.org/abs/2310.02304>.