# Philosopher Harry Frankfurt’s concept of bullshit now echoes in AI’s ‘botshit’ challenge



Lies have often been viewed as the greatest threats to truth, but philosopher Harry Frankfurt offers a more nuanced perspective, suggesting that the more insidious danger lies in what he calls "bullshit." In his seminal essay, *On Bullshit*, Frankfurt elucidates that while liars actively engage with truth by defying it, those who disseminate bullshit disregard truth altogether. This concept finds eerie resonance in the realm of generative artificial intelligence (AI), particularly within large language models (LLMs) such as ChatGPT and Claude.

Frankfurt passed away in 2023, shortly after the launch of ChatGPT, a development that prompts a reflection on his ideas in the context of modern technology. The outputs of these AI systems, which often produce plausible-sounding text without any grounding in factual accuracy, are being described as a form of "botshit" by scholars like Carl Bergstrom and Jevin West from the University of Washington. Their online course, titled *Modern-Day Oracles or Bullshit Machines?*, examines the challenges posed by such technologies. The models are adept at creating content that may appear authoritative yet lacks any substantiated truth, raising concerns about their impact on public discourse.

A particular concern is the phenomenon of "hallucination," where AI systems invent facts outright. Some researchers argue this may be an intrinsic characteristic of probabilistic models rather than a fixable flaw. Despite the efforts of tech companies to enhance AI reliability through improved data and fact-checking methodologies, the challenges remain significant. In a recent legal case, a lawyer from Anthropic admitted to inadvertently submitting a fabricated citation generated by the company’s AI. Such incidents highlight the real-world consequences of relying on AI for accurate data.

Google’s recent efforts to integrate AI capabilities into all its main services reflect a broader trend amongst tech giants. Their chatbot, Gemini, includes disclaimers about potential inaccuracies, yet this has not deterred its rollout. Experts have expressed concerns that improvements—such as reinforcement learning from human feedback—may inadvertently incorporate biases and subjective judgments, further complicating the quest for truth in AI outputs.

Moreover, the definition of “careless speech,” as articulated by researchers from the Oxford Internet Institute, underscores an alarming dimension of these technologies. This form of communication can inflict long-term, pervasive harm—akin to "invisible bullshit" that progressively diminishes societal understanding. In a landscape where human communicators typically have motivations that can be identified, AI chatbots operate without intentionality. They can fabricate information with no purpose other than generating engaging responses, which poses severe risks to the integrity of shared knowledge.

As conversations about the potential for more truthful AI models gain traction, it raises critical questions about market demand and whether developers should adhere to standards akin to those expected from professionals like advertisers or medical practitioners. Sandra Wachter, an academic in the field, likens the development of more reliable systems to the impracticality of turning a car into an aircraft—acknowledging that significant time, investment, and reformation in design philosophy would be essential.

Despite these concerns, generative AI continues to offer substantial utility across various sectors. Individuals and businesses are already harnessing its capabilities for innovation. However, conflating these models with reliable truth sources remains an illusion fraught with peril. The tech industry's rush towards implementation must balance enthusiasm for AI's transformative potential with a sober acknowledgment of its limitations and risks.

In conclusion, as the integration of AI into everyday life deepens, it becomes crucial to approach these systems with a critical eye. While they have the potential for enriching human productivity and creativity, understanding their nature as generators of plausibility rather than truth is essential to safeguarding public trust and maintaining the integrity of information.

**Reference Map**1: Paragraph 1, 2, 3, 4, 5, 6, 7
2: Paragraph 4, 5, 6
3: Paragraph 5, 6
4: Paragraph 7
5: Paragraph 5, 7
6: Paragraph 5, 6
7: Paragraph 5, 6

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.ft.com/content/55c08fc8-2f0b-4233-b1c6-c1e19d99990f> - Please view link - unable to able to access data
2. <https://www.ft.com/content/55c08fc8-2f0b-4233-b1c6-c1e19d99990f> - This article explores the parallels between generative AI models and the concept of 'bullshit' as defined by philosopher Harry Frankfurt. Unlike liars who engage with truth by rejecting it, bullshitters disregard truth entirely in favor of persuasion. Modern AI language models, such as ChatGPT and Claude, exemplify this dynamic by generating plausible-sounding content regardless of factual accuracy—termed 'hallucinations.' Professors and researchers have likened their outputs to 'botshit' or 'careless speech,' highlighting the risk of spreading misinformation without intentional deceit. Attempts to improve AI truthfulness—through better data, model tuning, and reinforcement learning—introduce their own challenges, including bias and subjective value judgments. As illustrated by a real court case involving fabricated AI-generated legal citations, these models’ unreliability can have serious consequences. Despite this, tech giants continue to expand AI integration across services. Experts suggest the creation of more truthful models will require significant investment, raising the question of whether market demand or regulatory standards will drive change. Ultimately, while generative AI offers utility across industries, mistaking it for a reliable source of truth is both naive and dangerous.
3. <https://www.ft.com/content/ed323f48-fe86-4d22-8151-eed15581c337> - This article discusses the immense enthusiasm surrounding generative AI, with major companies like Alphabet and Microsoft reshaping their businesses around it, and venture capitalists investing billions into AI startups. However, doubts exist regarding the reliability and practical utility of generative AI. Critic Gary Marcus highlights the technology's tendency towards inaccuracies or 'hallucinations,' and the potential issue of data pollution caused by AI-generated content. Despite these criticisms, supporters argue that generative AI can still enhance productivity, solve specific problems, and eventually create unimagined new services and business models. While the true economic value of generative AI remains uncertain, especially given the anticipated challenges, it has the potential to significantly transform various industries like past innovations did.
4. <https://time.com/6271657/a-to-z-of-artificial-intelligence/> - This comprehensive glossary provides explanations of terms related to artificial intelligence, including concepts like AGI (Artificial General Intelligence), alignment problems, automation, biases, chatbots, and competitive pressures. It also covers technical elements like neural networks, data requirements, compute power, and the differences between supervised, unsupervised, and reinforcement learning. The document discusses societal considerations such as regulation, redistribution of AI profits, lobbying, and the potential existential risks posed by advanced AI systems. The glossary aims to make AI terminology accessible to both beginners and those familiar with the field, enhancing public understanding of AI's power, promises, and perils.
5. <https://www.theatlantic.com/technology/archive/2024/08/chatbots-false-memories/679660/?utm_source=apple_news> - As AI chatbots and generative AI become more integrated into everyday tools like search engines and social media, the risk of misinformation increases. Research indicates that AI-generated responses can be highly persuasive, leading people to believe and internalize false information. Studies have shown that chatbots can implant false memories and distort reality, much like human manipulators. This potential for AI to subtly influence and mislead is particularly concerning as people use these tools to gather health information and make voting decisions. Tech companies claim they are working to ensure accuracy, but the inherent persuasive nature of AI outputs remains a significant challenge. Thus, the integration of AI into mainstream tools requires close scrutiny to prevent the spread of misinformation and protect public trust.
6. <https://apnews.com/article/82bc20f207e3e4cf81abc6a5d9e6b23a> - As major technology companies race to develop AI chatbots, glitches and inaccuracies persist. The enthusiasm around ChatGPT by OpenAI has prompted Google, Baidu, and Microsoft to announce their own versions, such as Google's Bard and Microsoft's Bing chatbot. While these chatbots promise current information and better integration with web searches, they often produce errors and inaccuracies, such as Bing's false Super Bowl narrative. Despite the imperfections, tech giants like Microsoft continue to push innovation, banking on AI as the future of the internet. These chatbots aim to improve search experiences by providing more engaging and immediate responses, though companies are cautious about making them appear too human.
7. <https://www.ft.com/content/741f905c-8aa7-4f72-a7f5-7afc19366e43> - The article discusses the similarities and differences between human memory and generative AI models, particularly focusing on the concept of hallucinations. Both humans and AI can misremember or produce incorrect information. Notable human examples include the discrepancies found in John Dean's Watergate testimony when compared to actual recordings. AI models exhibit similar issues, such as ChatGPT generating nonexistent cases or incorrect facts. Recent advancements, like the Safe methodology from Google DeepMind, aim to improve AI's factual accuracy by cross-referencing answers with Google Search. Maria Schnell emphasizes the importance of not just factual accuracy but also the relevance and contextual understanding tailored to specific audiences, a task still challenging for AI. The article concludes that while AI and humans have their flaws, collaboration between the two can leverage their strengths and reduce errors.