

# The Top 10 AI Myths of 2023

Myth: *“AI is unbiased.”*

Fact: “AI is trained on human data saturated with 200+ cognitive biases, and it operates based on heuristics, giving it the dynamics of human cognitive bias. It also lacks any factual grounding, making any illusions of unbiased assessment both superficial and transient.”

**Myth: *“AI-generated text and images can be detected.”***

**Fact: “All detectors designed for AI-generated text and images are vulnerable to adversarial tactics, and most never worked to begin with. Text and images can be slightly and automatically modified to bypass detection.”**

**Myth: *“LLMs can be used to create intelligent search engines.”***

**Fact: “LLMs aren't data compression or retrieval systems, nor can the misinformation they generate be reliably detected. Websites can be designed to poison, corrupt, and take over LLMs via Indirect Prompt Injection, making them not only untrustworthy but also insecure. These security vulnerabilities also grow rapidly with each integrated app and ‘multi-modal’ capacities.”**

***Myth: “LLMs and similar AI can ‘unlearn’ information to comply with legal requirements.”***

**Fact: “Since LLMs aren’t compression or retrieval systems they never truly learn information, nor can the heuristics they train be removed without deleting the model. Even if all data about an individual were hypothetically removed, a blank space in a giant auto-complete function will always be filled, and what it is filled with may be worse than what was there before.”**

**Myth: *“There are many AI Experts working on human-like AI.”***

**Fact: “Expertise in LLMs, RL, and other conventional narrow AI confers precisely zero expertise into human-like or human-level intelligence or understanding.**

**Anthropomorphism is rampant, with abundant (false) comparisons made to the human brain.”**

***Myth: “AI guardrails can be developed and deployed, adding safety and security to LLMs and integrated systems.”***

**Fact: “Every single attempt at guardrails has been torn to shreds by Security Researchers, often within 5 minutes, on the first attempt, and more recently with automated adversarial systems that were able to easily break every single LLMs guardrails, even when they were only trained against two open-source models. Guardrail methods also degrade the performance and usability of models, without exception. Any system that can be “prompt engineered” can also be “prompt injected”, as the two are functionally identical. Companies proposing these are committing fraud in no uncertain terms.”**

**Myth: *“Society can handle a bit of AI-generated misinformation, it isn’t a problem.”***

**Fact: “Most AI-generated misinformation is never detected. Since that misinformation can’t be reliably detected as AI-generated it is also fed back into training new LLMs by companies scraping the internet looking for more data. This causes the new LLMs to be poisoned by quickly increasing amounts of misinformation, in a very hazardous negative feedback loop. No system yet exists to counter this growing threat.”**



**Myth: *“Conventional AI like LLMs can be used to assist in research.”***

**Fact: “Growing trends among researchers, such as using GPT-4 to grade its own performance, and the performance of competing models, are tantamount to fraud. These abuses of the technology cannot be called ‘research’ or ‘scientific’ by any stretch of the imagination. While LLMs may, in theory, be used in some helpful capacity, this manner of fraud is the current reality.”**

**Myth: *“Foundation Models demonstrate emergent capacities, and AGI may emerge from them.”***

**Fact: “The ‘emergent capacities’ in conventional AI were thoroughly debunked by researchers at Stanford earlier this year. There is nothing magical or spontaneous that occurs in these models. ‘Foundation models’ also aren’t a foundation for AGI, as they offer none of the required capacities, they are just chatbots.”**

***Myth: “If we keep working on these problems, LLMs can become Safe, Ethical, Trustworthy, Aligned, and Explainable.”***

**Fact: “The fundamental architectural limitations of LLMs and other conventional AI cannot be overcome with any amount of time or effort. The architectural choices that went into their design are incompatible with these concepts. Major architectural changes and additions, which none of the major tech companies and startups are currently working on, are required to make any meaningful progress on addressing these issues. Again, companies perpetuating this myth are committing fraud in no uncertain terms.”**

**Thank you for choosing to learn!**

**Humanity needs more people to choose reality over fantasy.**

**Cognitive biases offer a constant temptation to impoverish our decision-making and grasp of reality. Though humans aren't built to fully overcome these biases, we may reduce their ability to influence us.**

*“One of the biggest problems with the world today is that we have large groups of people who will accept whatever they hear on the grapevine, just because it suits their worldview—not because it is actually true or because they have evidence to support it. The really striking thing is that it would not take much effort to establish validity in most of these cases... but people prefer reassurance to research.” — Neil deGrasse Tyson*