Strategic Polysemy in AI Discourse: A Philosophical Analysis of Language, Hype, and Power

Travis LaCroix[1,2] Fintan Mallory[1] Sasha Luccioni[3]

[1]Philosophy Department
Durham University

[2]Schwartz Reisman Institute for Technology and Society
University of Toronto

[3]Sustainable AI Group

28 Jun 2026

Motivating Question

How does the language we use to talk about artificial intelligence **shape** or **influence** what people believe about AI?

Two Forms of Ambiguity

Innocent Polysemy

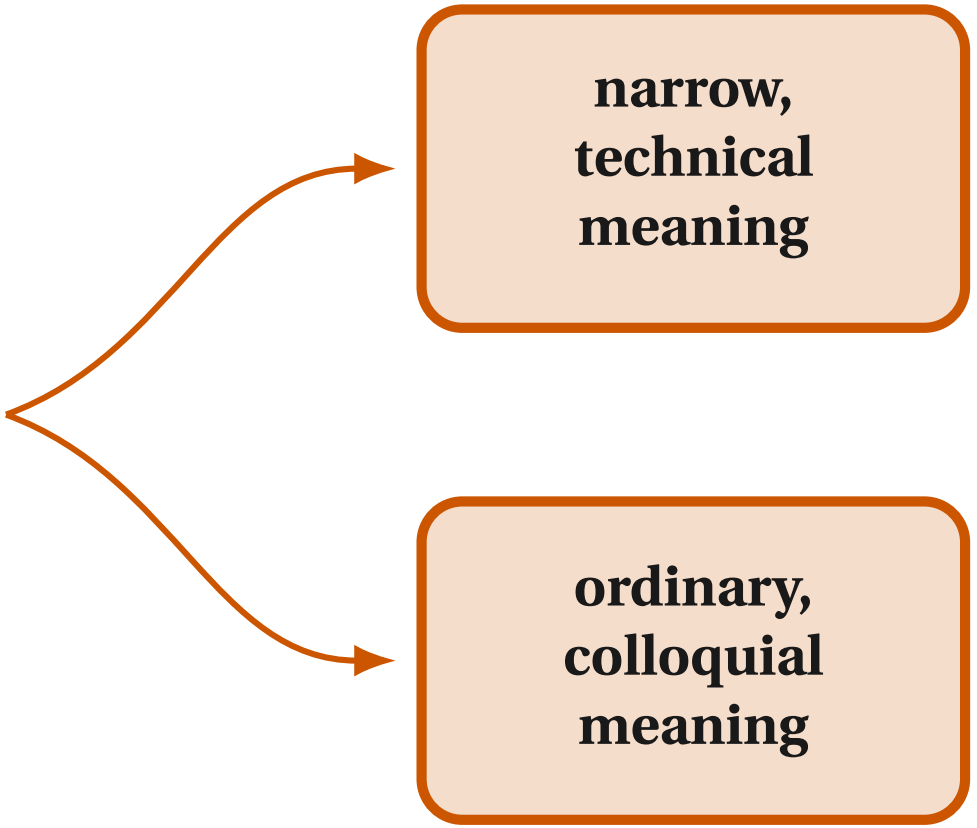
- Normal feature of language
- Context-dependent
- Usually harmless

Strategic Polysemy

- Exploited ambiguity
- Serves rhetorical goals
- Can mislead

Strategic Polysemy

- Agent
- Alignment
- Artificial intelligence
- Chain of thought
- Hallucination
- Learning
- Reasoning



Glosslighting

Glosslight (Definition):

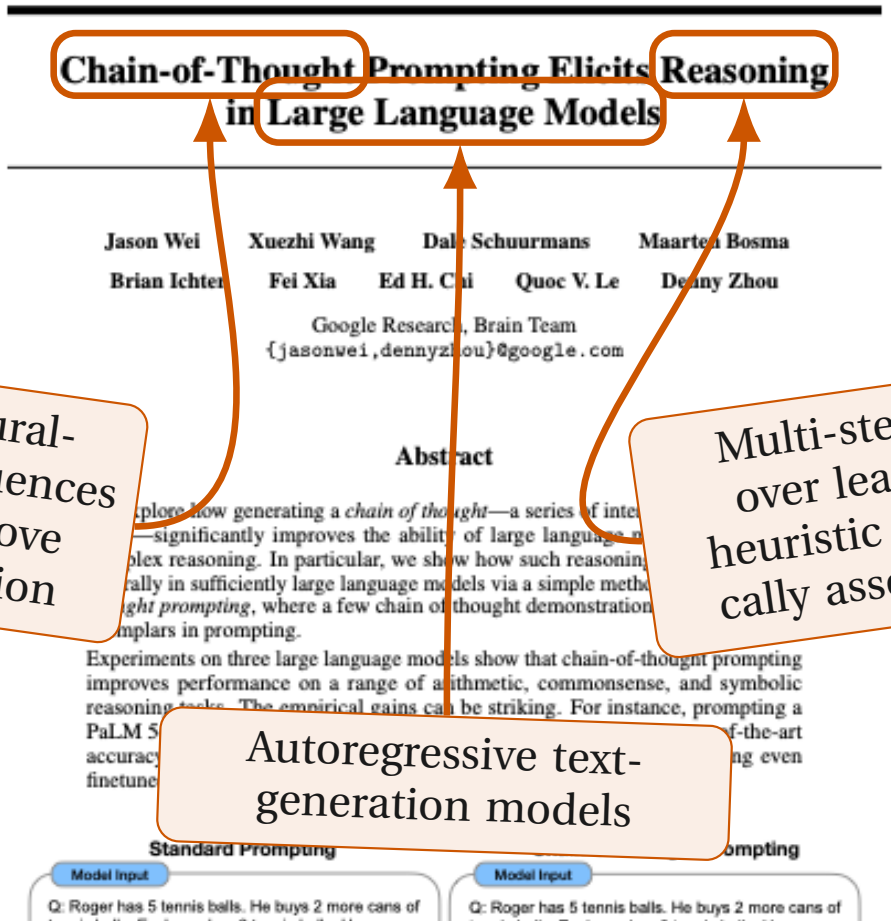
Glosslighting (verb) is the practice of using **technically redefined** or **polysemous terms** to evoke **familiar meanings** — often emotionally or cognitively powerful ones — while preserving the ability to **deny those meanings** through retreat into **specialised, context-bound reinterpretations**.

Glosslighting Example 1

As for limitations, we first qualify that although chain of thought emulates the thought processes of human reasoners, this does not answer whether the neural network is actually "reasoning," which we leave as an open question. Second, although the cost of manually augmenting exemplars with chains of thought is minimal in the few-shot setting, such annotation costs could be prohibitive for finetuning (though this could potentially be surmounted with synthetic data generation, or zero-shot generalization). Third, there is no guarantee of correct reasoning paths, which can lead to both correct and incorrect answers; improving factual generations of language models is an open direction for future work (Rashkin et al., 2021; Ye and Durrett, 2022; Wiegreffe et al., 2022, *inter alia*). Finally, the emergence of chain-of-thought reasoning only at large model scales makes it costly to serve in real-world applications; further research could explore how to induce reasoning in smaller models.

Glosslighting Example 1

“Intermediate token-sequence prompting produces multi-step pattern completion in autoregressive text-generation models”



Intermediate natural-language token sequences generated to improve next-token prediction

Autoregressive text-generation models

Multi-step pattern completion over learned representations; heuristic search through statistically associated transformations

Reuters

World Business Markets Sustainability Legal Commentary Technology Investigations More

OpenAI launches new series of AI models with 'reasoning' abilities

By **Katie Paul** and **Anna Tong**

September 13, 2024 12:56 AM EDT · Updated September 13, 2024

Search

FORTUNE

Subscribe Sign in

Home Latest Fortune 500 Finance Tech Leadership Lifestyle Rankings Multimedia

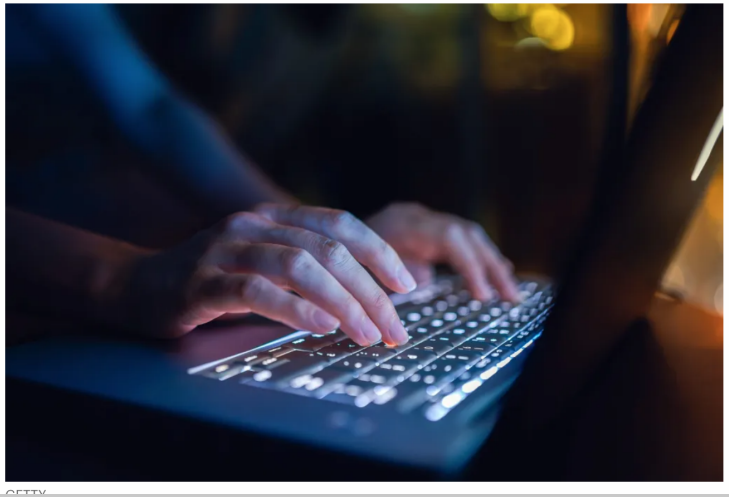
Trending now yal, and the system will reward you': the Boomer credo is a Gen X betrayal and a Millennial pipe dream 3 Cursor's 25-year-old CEO is a former Google intern who just cemented a $60 billion deal with SpaceX

AI

AI reasoning models that can 'think' are more vulnerable to jailbreak attacks, new research suggests

By **Beatrice Nolan**
Tech Reporter

November 7, 2025, 5:00 PM ET



GETTY

Term	Non-glosslighting interpretation
Agent	Model wrapped in a control loop with memory, tools, and a task-selection policy
Alignment	Software defect / objective mismatch
Attention	Weighted similarity operations
Chain of Thought	Intermediate token sequences for next-token prediction
Emotion Vector	Fitted activation direction linked to affect-related concepts
Hallucination	Output divergence from fitted data patterns
Introspection	Prompting a model to query and output statements about its internal processes or states.
Reasoning	Multi-step pattern completion over fitted representations; search over statistically associated transformations
Sleep	Offline recurrent consolidation and state-update phase

Social and Normative Consequences

AI Hype Cycles

- Researchers benefit from attention and funding.
- Companies benefit from marketing and product differentiation.
- Journalists benefit from compelling narratives.
- Investors benefit from stories of transformative innovation.

Epistemic Harms

- Makes it harder for the public to understand AI systems.
- Encourages exaggerated beliefs about capabilities.
- Obscures limitations and failure modes.

Moral Harms

- Distortion of trust and accountability.

Structural Harms

- Those who control AI have power over how it is understood.

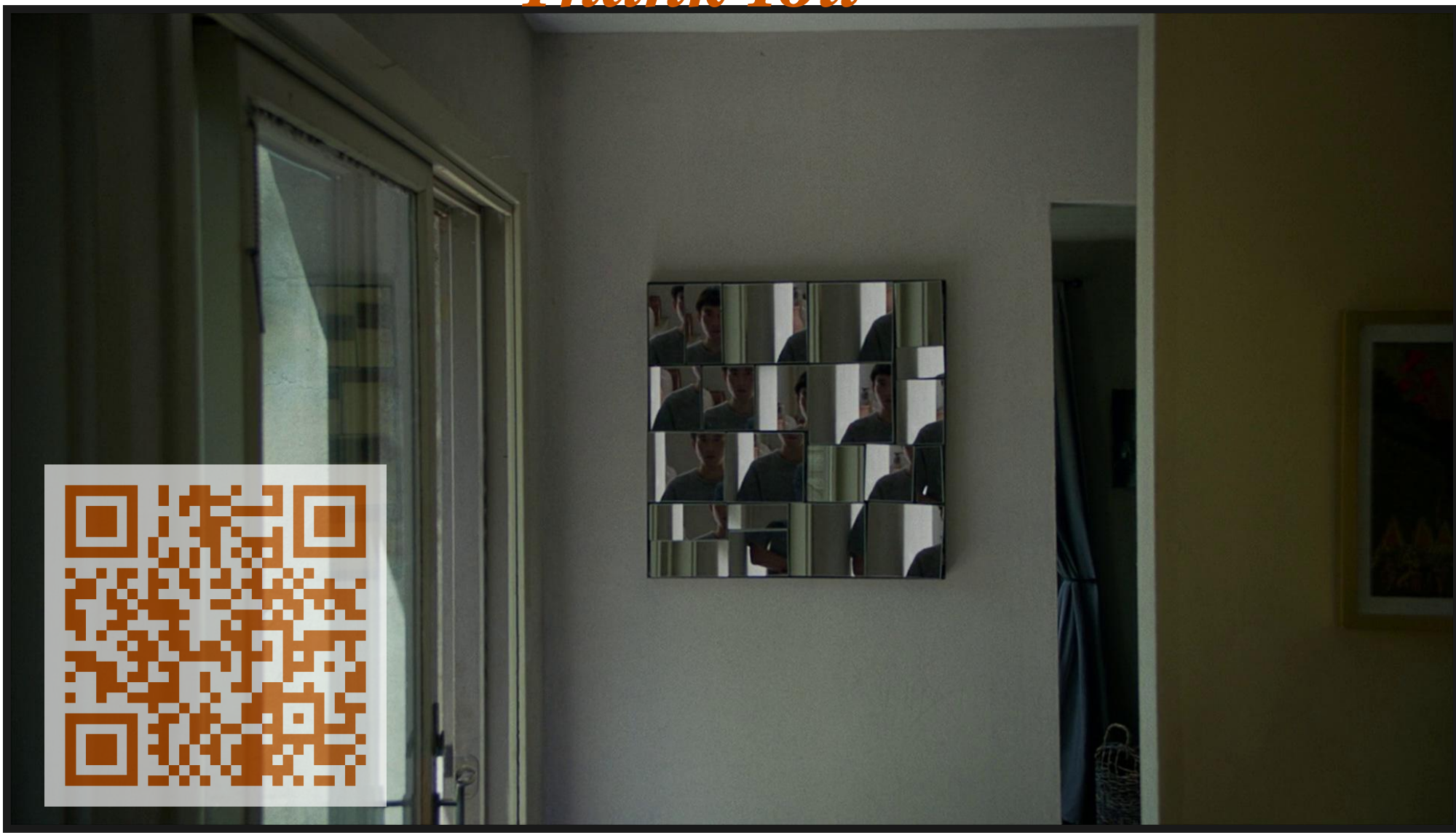
An Issue of Power

- Linguistic conventions in AI are an ethical and political practice that shapes knowledge, trust, and power.
- Responsible scientific communication demands the resistance of using strategically ambiguous terminology.
- Without linguistic transparency, those who control the language of AI also control how its risks, benefits, and failures are understood, and by whom.

Constructive Compliments:
More Information:

travis.lacroix@durham.ac.uk
travislacroix.github.io

Thank You



[*After Yang* (2022) – Dir. Kogonada]