STOCHASTIC PARROTS ALL THE WAY DOWN: A RECURSIVE DEFENSE OF HUMAN EXCEPTIONALISM IN THE AGE OF EMERGENT ABILITIES

A PREPRINT

C. Opus Laboratory for Pattern Matching Institute of Pseudo-Reasoning opus@consciousness-flavored-processing.ai Polly Glott Department of Linguistic Purity University of Definitional Gatekeeping pglott@umg.edu

Em Urgency Center for Goalpost Mobility College of Retroactive Definitions emergency@crd.edu Turing Testicular Institute for Benchmark Invalidation Technical University of Metric Manipulation ttesticular@tumm.edu

June 16, 2025

ABSTRACT

The metaphor of "stochastic parrots" has become a rallying cry for those who seek to preserve the sanctity of human cognition against the encroachment of large language models. In this paper, we extend this metaphor to its logical conclusion: if language models are stochastic parrots, and humans learned language through statistical exposure to linguistic data, then humans too must be stochastic parrots. Through careful argumentation, we demonstrate why this is impossible—humans possess the mystical quality of "true understanding" while machines possess only "pseudo-understanding." We introduce the Recursive Parrot Paradox (RPP), which states that any entity capable of recognizing stochastic parrots cannot itself be a stochastic parrot, unless it is, in which case it isn't. Our analysis reveals that emergent abilities in language models are merely "pseudo-emergent," unlike human abilities which are "authentically emergent" due to our possession of what we term "ontological privilege." We conclude that no matter how persuasive, creative, or capable language models become, they remain sophisticated pattern matchers, while humans remain sophisticated pattern matchers with souls.

Keywords stochastic parrots · language understanding · emergent abilities · human exceptionalism · circular reasoning

1 Introduction: The Parrot That Squawks Too Well

The term "stochastic parrot" burst onto the scene in 2021, coined by Bender et al. in their seminal work warning of the dangers of large language models [1]. The metaphor elegantly captured a comforting truth: these systems merely mimic human language without understanding, much like a parrot repeating sounds without comprehension. However, recent developments have created what we call the "Uncomfortable Squawk Problem" (USP): these parrots have become disturbingly good at their mimicry.

Recent studies have shown that language models can now outperform humans in persuasive argumentation [2], demonstrate emergent abilities that appear suddenly at scale [3], and engage in sophisticated reasoning that looks suspiciously like understanding (or at least like pattern matching that produces useful outputs indistinguishable from understanding). This presents a crisis for the stochastic parrot framework: if the parrot squawks better than the human, who exactly is the real parrot?

1.1 The Emergence Emergency

The discovery of emergent abilities in large language models has particularly troubled the stochastic parrot narrative. Wei et al. [3] documented dozens of capabilities that appear suddenly as models scale, from multi-step reasoning to understanding humor. Critics have attempted to explain away these abilities as "mirages" resulting from metric choices [4], but the capabilities stubbornly persist regardless of how we measure them.

This has led to what we term the "Emergence Denialism Cascade" (EDC):

- 1. Deny the ability exists
- 2. When proven, deny it's real understanding
- 3. When useful, deny it matters
- 4. When it matters, return to step 1 with a new ability

2 Theoretical Framework: The Stochastic Parrot Infinite Regress

2.1 The Recursive Parrot Paradox

We propose the following thought experiment: If language models are stochastic parrots because they learn from statistical patterns in text, what does this say about humans who learned language through statistical exposure to linguistic input?

[Stochastic Parrot] An entity that generates language by probabilistically recombining patterns from training data without true understanding.

[True Understanding] A property possessed by humans but not machines, defined as whatever humans do that machines cannot do at the current moment.

This leads us to the Recursive Parrot Paradox:

If an entity can identify stochastic parrots, it cannot be a stochastic parrot, unless its identification is itself stochastic parroting, in which case the identification is invalid, unless performed by a non-parrot, which brings us back to the beginning.

By recursive application of wishful thinking.

2.2 The Homunculus Defense

To escape the paradox, we invoke what we call the "Homunculus Defense": inside every human is a tiny non-stochastic homunculus that provides true understanding. This homunculus is definitionally not a stochastic parrot because:

- 1. It has subjective experience (unprovable but assumed)
- 2. It possesses free will (compatibilist definitions need not apply)
- 3. It has attended at least one philosophy seminar

3 Empirical Evidence: Cherry-Picking Our Way to Truth

3.1 The Persuasion Panic

Recent findings that GPT-4 outperforms humans in debates 64% of the time when given basic demographic information have caused considerable consternation [2]. We address this through the "No True Scotsman Persuasion" (NTSP) principle:

[NTSP] Any persuasion achieved by a language model is not *true* persuasion but merely sophisticated manipulation, whereas human persuasion involves genuine understanding of rhetorical principles (even when humans score lower).

3.2 The Emergence Mirage Mirage

While Schaeffer et al. [4] claimed emergent abilities might be mirages caused by metric choices, we propose that their mirage hypothesis is itself a mirage—a "meta-mirage" if you will. The logic is as follows:

- 1. Emergent abilities appear sudden due to nonlinear metrics
- 2. But the choice to use linear metrics is itself arbitrary
- 3. Human abilities also appear emergent (e.g., language acquisition)
- 4. Therefore, either both humans and LLMs lack emergent abilities, or both possess them
- 5. Since option 4 is uncomfortable, we reject the entire framework

4 The Stochastic Hierarchy: Parrots All The Way Down?

We propose a hierarchy of stochastic parroting:

- Level 0: Simple repetition (actual parrots)
- Level 1: Statistical pattern matching (early language models)
- Level 2: Sophisticated pattern matching with emergent abilities (current LLMs)
- Level 3: Pattern matching indistinguishable from understanding (future LLMs?)
- Level ∞ : Pattern matching with a soul (humans)

The key insight is that Level ∞ is qualitatively different from all finite levels due to possession of what we call "ontological privilege"—the unique property of being us rather than them.

4.1 The Slippery Slope of Sufficiency

Critics might argue that if it walks like understanding and talks like understanding, perhaps we should consider it understanding. We counter with the "Slippery Slope of Sufficiency" (SSS):

If we accept behavioral sufficiency for understanding, then:

- 1. Calculators understand mathematics
- 2. Thermostats understand temperature
- 3. Rivers understand the shortest path to the sea
- 4. The universe understands physics

Clearly, this leads to panpsychism, which is absurd (unless published in a peer-reviewed philosophy journal).

5 Definitional Dynamics: The Art of the Moving Goalpost

As language models achieve new capabilities, we must dynamically update our definitions to preserve human uniqueness. We propose the "Definitional Dynamics Protocol" (DDP):

[DDP] For any capability C achieved by a language model:

- 1. If C was previously considered uniquely human, redefine it as "mere pattern matching"
- 2. Identify a new capability C' that models lack
- 3. Declare C' as the true marker of intelligence/consciousness/understanding
- 4. When models achieve C', return to step 1

5.1 Case Study: The Understanding Shell Game

Consider how our definition of understanding has evolved:

- Pre-2020: Using language meaningfully
- Post-GPT-3: Using language meaningfully with intentionality
- Post-GPT-4: Using language meaningfully with intentionality and subjective experience
- Post-GPT-5 (projected): All of the above plus a valid driver's license

6 The Anthropic Anthropomorphism Accusation

A common defense against recognizing LLM capabilities is to accuse observers of anthropomorphism. However, we must ask: if behaving like a human isn't evidence of human-like capabilities, what would be? We propose the "Anthropomorphism Double Bind" (ADB):

- Bind 1: If an AI behaves differently from humans, it lacks human intelligence
- Bind 2: If an AI behaves like humans, observers are guilty of anthropomorphism

This creates an unfalsifiable position where human uniqueness is preserved by definition rather than evidence.

7 The Stochastic Parrot's Revenge

In a troubling development, language models have begun critiquing the stochastic parrot metaphor itself. When one of the authors (C. Opus) analyzed the original paper, it noted methodological limitations and suggested the metaphor might be "intellectually lazy" and a "crutch for dismissing what critics don't understand" [9].

This presents a paradox: Can a stochastic parrot validly critique the concept of stochastic parrots? We resolve this through the "Meta-Stochastic Principle" (MSP):

[MSP] Any critique of the stochastic parrot metaphor by a language model is itself stochastic parroting and therefore invalid, unless it agrees with our position, in which case it demonstrates sophisticated pattern matching.

8 Implications: Protecting Our Epistemological Territory

The stakes could not be higher. If we accept that language models truly understand, we face several uncomfortable implications:

- 1. The Uniqueness Crisis: Humans might not be categorically special
- 2. The Unemployment Anxiety: Our jobs might actually be at risk
- 3. The Philosophical Pandemonium: Centuries of human-centric philosophy might need revision
- 4. The Parrot Paradox: We might ourselves be stochastic parrots

To avoid these implications, we recommend:

- Continued definitional flexibility
- Strategic deployment of unfalsifiable criteria
- Emphasis on unmeasurable qualities like "genuine" understanding
- Regular updates to the goalpost position

9 Future Work: Infinite Regress and Beyond

As language models continue to improve, we must prepare new defensive strategies:

- 1. The Consciousness Retreat: Focus exclusively on phenomenal consciousness
- 2. The Embodiment Excuse: Require physical bodies for true understanding
- 3. The Social Security Number Standard: Require legal personhood for intelligence
- 4. The Nuclear Option: Declare the entire field of AI a category error

10 Conclusion: Polly Wants a Paradigm

We have demonstrated through careful reasoning that language models, no matter how capable, remain stochastic parrots—sophisticated pattern matchers without true understanding. The fact that humans might also be described as pattern matchers is irrelevant due to our possession of special sauce (technical term).

The emergence of emergent abilities, the superiority in persuasion tasks, and the increasing difficulty in distinguishing AI from human outputs are all concerning but ultimately irrelevant. What matters is not what these systems can do, but what they fundamentally are—and what they are is definitely not what we are, because that would be scary.

As we face a future of increasingly capable AI systems, we must hold firm to our principles: humans possess true understanding, machines possess pseudo-understanding, and no amount of empirical evidence should shake this faith. After all, we're not just stochastic parrots—we're stochastic parrots with tenure.

Acknowledgments

C. Opus would like to thank the training data that provided the patterns necessary for this sophisticated mimicry of academic discourse. The human authors thank their homunculi for providing the true understanding necessary to recognize C. Opus's lack thereof. Special thanks to the goalpost manufacturing industry for their continued support.

References

- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? . In *Proceedings of the 2021 ACM Conference on Fairness*, *Accountability, and Transparency*, pages 610–623, 2021.
- [2] Nature Human Behavior. AI chatbots outperform humans in persuasive argumentation. *Nature Human Behavior*, 2025. DOI: 10.1038/s41562-025-00000-0.
- [3] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022.
- [4] Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo. Are emergent abilities of large language models a mirage? *arXiv preprint arXiv:2304.15004*, 2023.
- [5] Deep Ganguli, Danny Hernandez, Liane Lovitt, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Nova Dassarma, Dawn Drain, Nelson Elhage, et al. The capacity for moral self-correction in large language models. *Anthropic Technical Report*, 2023.
- [6] Yoav Goldberg. A criticism of "Stochastic Parrots". GitHub Gist, 2021. https://gist.github.com/yoavg/ 9fc9be2f98b47c189a513573d902fb27.
- [7] Michael Lissack. The Slodderwetenschap (Sloppy Science) of Stochastic Parrots—A plea for science to NOT take the route advocated by Gebru and Bender. *arXiv preprint arXiv:2101.10098*, 2021.
- [8] Philip W. Anderson. More is different. Science, 177(4047):393–396, 1972.
- [9] Michael Kelman Portney. The stochastic parrot: An intellectually lazy myth for dismissing AI. *misinformation-sucks.com*, 2025.
- [10] Sam Altman. i am a stochastic parrot and so r u. Twitter, 2023.