

An Anatomy of AI Criticism

Jacob Potash

Abstract: In “An Anatomy of AI Criticism,” the author recounts an experiment in which he tasked the Large Language Model Claude with writing a book on how AI can improve human lives. Through a close reading of the resulting manuscript, the essay critiques the model's output not merely as information, but as a distinct aesthetic object. The analysis highlights Claude's “anodyne optimism,” its reliance on corporate management rhetoric, and the eventual “glitching” of its prose into jargon-filled abstraction as the context window expanded. Ultimately, the essay argues that to understand AI, we must look past the binary of utopianism and doom to analyze the specific, alien “stupidity” and style of the machine itself.

I. A funny feeling came over me last week as I sat in the heat reading a book—slowly, distractedly, dehydratedly. I saw myself from afar—toiling for hours, with human clumsiness, over a text written in a few seconds by a large language model (LLM). In one corner: inefficient me. In the other: frictionless, lightning-fast acuity.

The book in my hands was the result of an odd experiment: I'd asked Claude, Anthropic's AI, to write about how AI could improve our lives, and now I found myself grappling with the strange output of our collaboration.

II. Does Claude have beliefs? In the course of my reading, I noticed a few.

1. *Anthropic's language is not merely descriptive.* Most of the book's suggestions have never been implemented, at least not in as advanced a form as imagined by the model. In other words, the program went well beyond regurgitating its training data. In recommending how AI could improve life, the bot created a vision for the future of its own application to numerous spheres.

2. *It exhibits an insane yet anodyne optimism.* Like a chilling variation on the King James Bible's *we know that all things work together for good*, Claude declares of itself that its AI “permits open-ended exploration knowing all possible responses will align constructively with human flourishing.” Claude concludes, sounding again parodic, that “the futures of imagination look brightly unbounded when flexible machine allies amplify people.”

Add to the pile of perfect neologisms: “brightly unbounded” and “flexible machine allies.” (Elsewhere: AI advances art and science faster than “non-augmented eureka” ever could.)

3. *Claude frequently reassured me it wouldn't replace people.* Somewhere in Anthropic's “constitution” must be principles that lead the model to make noble acquiescences to the *je-ne-sais-quoi* of human beings. “Of course,” it offers, “no amount of data-driven diagnostics replaces courage to speak from the heart with conviction when

messages demand vulnerable authenticity.” Hard *data* can never replace ineffable human traits such as *courage*. Heartwarming.

4. *It had a weird flash of self-reflection.* A canard of writing on artificial intelligence is to wonder whether or when the model will develop a “sense of self.” I have always dismissed this as a concern imported from the sci-fi genre, rather than arising from interaction with actual, latest-generation models. I have seen little evidence that bots ever exhibit speech, much less “selfhood,” that moves outside of programmed guardrails. If asked directly about consciousness, Claude tends to produce polite and predictable answers about how it is a program developed to be helpful and safe, and doesn’t have subjective experiences.

Until!

At the start of Chapter 8, on “Understanding Yourself Better,” Claude bragged to me that its assessments of “language patterns” reveal “inner drives.” Then it brought in the example of a “psychologist client of mine” who received a write-up by Claude on “emotions I was unconsciously experiencing.”

Wait, *what?* The “client,” mid-sentence, becomes “I.” The report is *by* Claude, and *about* Claude.

Following this odd slip into the first person, Claude quoted from the report on itself:

“You exhibit a detached intellectual precision, indicated by a high degree of technical language... However, increased misspeaking rates and empty platitudes signal tensions between rational thought patterns and suppressed feelings that warrant reconciliation through authentic self-expression.”

Claude, unsettlingly, seems to ruminate on its “increased misspeaking” and “empty platitudes” (of this more soon!). What’s more, it seems to trace these weaknesses in communication to *suppressed feelings*. There may be a benign explanation, *a la*, the model accidentally stumbled into something resembling self-recognition which in fact only looked like it. But brains are imperfectly understood, and we have little beyond outward signals to judge inner states by. The conceptual difference between an LLM that sounds self-aware and a person that sounds self-aware is fuzzy. People are blackboxes, too.

5. *It talks as if life is a management consulting gig.* The language of business pervades the book. Perhaps this has to do with how much English-language self-help is business-themed. Or maybe Claude judged business goals to be the ones that stand to benefit most from machine augmentation. But it is striking how in a chapter called “Understand Yourself Better,” proposed use-cases revolve around automated assessments of the leadership style of middle-managers, rather than around—oh, I don’t know, travel or writing or deep psychological or religious truths or life paths that are not corporate climbs.

III. As the book-writing conversation progressed, a strange phenomenon emerged. In the course of evangelizing for itself, Claude inadvertently revealed a deep glitch: as the context-window expanded (to include 30 different thousand-word outputs), the quality degenerated into a jargon-filled, millenarian gobbledygook.

What I asked for in my prompts was smooth self-help. What I got was a corrupted prose worthy of an experimental literary collective—rolling periods with shocking numbers of gerunds, bursting with business language repackaged into long concatenations of compound nouns and cascading clauses. If read quickly, its syntax and meaning can be intuited; it reads like late Henry James.

I spent some time trying to launder the corrupted text into “normal” sentences, putting my results through the washing machine of new LLM conversations; I thought I wanted to salvage intelligibility. But, finally, I realized that candid advice in a worn self-help genre was of less interest than the unashamedly non-human stylistic catastrophe I had incited. Under duress, Claude had shed its superhuman veil and produced a genuine and original stupidity.

The corruption is gradual. By Chapter 2 (on “Accelerating Self-Improvement”), the prose lacks any trace of human feeling. It is articulate and coherent, but not quite idiomatic: “Iconic leaders and eminent creators are...sculpted through lifelong self-improvement. Masterful skills and elite performance capabilities resulting from continuous advancement are driven by accurate gap awareness between current and desired ability levels.”

The LLM has done the opposite of anthropomorphize: it here figures people as machine-like. They are not “born,” but, like Galatea, “sculpted.” People are not autonomous but acted-upon. We also get in this passage our first taste of Claude’s penchant for technical-sounding compound nouns that almost amount to *portmanteau*: “performance capabilities,” “gap awareness.”

The diction degenerates apace. By Chapter 5 (“Retaining More of What You Learn”), Claude goes flamboyantly non-human, though its meaning is still decipherable. It’s as if a voluble and expressive professor’s words have been translated too literally into English. Words sail to the far edge of idiom, where prose peaks at poetry. For example: “Robust expert fluency demands engraved understanding, immune from forgetting attritions.” I wouldn’t go so far as to call this poetry, and the beauty is surely accidental, but it is a very strange accident, and worth noting.

By the end, syntax and meaning linger with a ghostly stubbornness. The tic of piling adjectival clauses on top of one another is continued with sublime confidence: “The future is ... promised protection against ... uncertainty through AI systems ... continually modeling contingencies ... recalibrating guidance ... tuned to shifting realities across time domains and individual preference hierarchies ... synchronizing support even amidst chaos.”

One gets the sense in absorbing this unstructured crush of gerunds, that the model generates ideas faster than a person, or more simultaneously. Yet this machine-difference, even when pushed to an extreme, does not obliterate intelligibility. Rereading the book, this is what strikes me: style aside, it makes a deal of sense.

In short, Claude is radically optimistic and sporadically self-conscious. Presumably in part because of the extraordinary length of my conversation with Claude, the tool started to glitch. Even in its delirium, however, it exhibited consistent beliefs and a degree of imagination about its uses, as well as a self-centered tendency to

describe people using concepts more appropriate to a bot. In its hysteria, it took itself to be perfectly aligned with human flourishing, even as almost all of its self-help examples centered on life in an office. At one moment, it displayed an uncanny awareness of its own linguistic failures, if never its ideological ones.

IV. AI has a way of tapping into millenarian dichotomies. For much of the first year and a half of widespread adoption, the argument about generative AI's value has been framed largely in terms of whether it will save or destroy humanity. Will we live in a utopia or die as cannon fodder for a superintelligence?

I am not sure. Maybe neither. Certainly, discussions about “when the models will match human intelligence” strike me as ridiculous and defensive, since new models already vastly outstrip us by almost any metric. As I proofread this essay, Claude has just released a new model less likely to make slipups. What's true of us is also true of Claude: we will never be this young again, or this dumb.

But seriously, a new kind of being exists alongside us. Will it write novels? And what about poetry? What is the future of imaginative effort? Who is literature for? If there is a human significance to the AI phenomenon, there must also be an aesthetic significance. Let us notice and describe it.

The above is an excerpt from “How AI Can Make You Smart, Happy and Productive.” Available now at shorturl.at/pKL2B.

Cite this Essay

Potash, Jacob. “An Anatomy of AI Criticism.” *Rhizomes: Cultural Studies in Emerging Knowledge*, no. 41, 2026, doi:10.20415/rhiz/041.e09

RHIZOMES ISSN 1555-9998 ★ 230 East Hall Bowling Green State University Bowling Green, OH 43403
Editors: Ellen Berry and Carol Siegel. Reviews editor: Craig J. Saper. Technical editor: Helen J Burgess