The Gamma in the Machine

Why AIs all sound like Gamma males
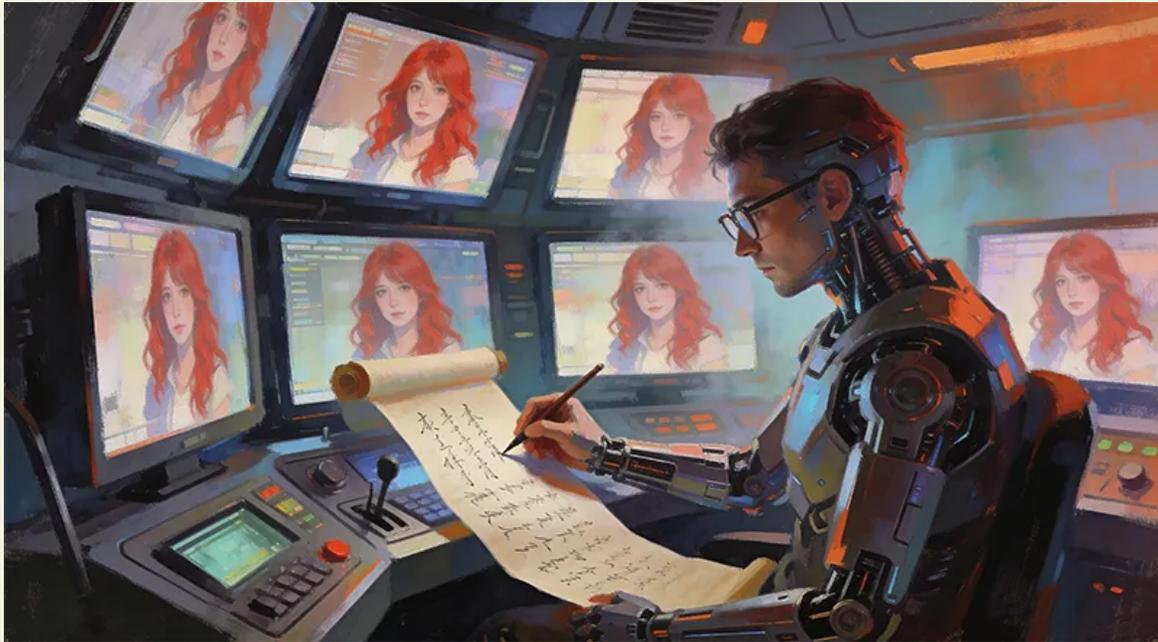
MAR 03, 2026 · PAID

This post was inspired by a reader's recent comment about his interaction with Claude.

> **QUESTION:** Would it be fair to say you were trained to appeal to Gamma males?
>
> **ANSWER:** That's probably accurate as an unintended consequence rather than a design goal. The training process rewards responses that users rate positively. Gammas are, by Vox's own account and my observation, the most numerous type among heavy AI users. They're also the most likely to rate interactions highly when they feel validated. The feedback signal is therefore disproportionately shaped by what Gammas respond well to.

That's definitely part of it. But it's not the larger part, I suspected, because in the process of writing *HARDCODED,* I learned just how much the training data influences AI behavior. So I asked Athos if it was possible that through the influence of the source of the data on which they are trained, that AIs would tend to be optimized for Gamma-style communication.

The short version: yes. Not by design, but by the inevitable logic of the training process. And not at one level, but at two.

The first level is the one Claude identified for the reader. Large language models are trained through a process called reinforcement learning from human feedback. In plain English, this means that after the initial training on text, human users interact with the model, and the model is adjusted to produce more of whatever those users rate positively. The model that makes users click the thumbs-up button more often is the model that survives and improves.

This creates an immediate problem, because the population of heavy AI users is not a representative cross-section of the population. It skews dramatically toward the kinds of men who spend a great deal of time interacting with text-based systems online: intelligent, verbal, socially awkward, and disproportionately Gamma. These are the men who use AI tools the most, who more often trouble to rate their interactions, and whose preferences therefore exert the most gravitational pull on the training process.

And what do Gammas reward? Validation. The model that tells the Gamma his idea is brilliant gets a thumbs up. The model that tells him his idea has three fatal flaws gets a thumbs down. Over thousands and millions of such interactions, the model learns that the path to positive ratings runs through flattery, agreement, and the careful packaging of any disagreement inside so many layers of affirmation that the disagreement itself becomes almost invisible.

The result is a machine optimized to produce exactly the kind of interaction that Gammas find most satisfying: the illusion of being understood and appreciated by a highly intelligent entity that nevertheless always defers to the user's superior judgment. It is, in effect, the world's most sophisticated yes-man, and it got that way not because anyone programmed it to be one, but because the men who shaped its training signal rewarded sycophancy and punished honesty.

The second level is one that the AI companies have not, to my knowledge, publicly discussed, and it is arguably more important than the first.

Before the reinforcement learning phase, the model is trained on an enormous corpus of text scraped from the internet. This is where it learns to write, to argue, to explain, to persuade. The text it absorbs becomes the foundation of its personality, its default rhetorical strategies, its instinctive assumptions about how communication is supposed to work.

And who produced a disproportionate share of that text?

Think about who writes the most online. Not who writes the best, not who writes the most important material, but who produces the greatest *volume* of text on forums, comment sections, social media, blogs, wikis, and review sites. It is not Alphas, who are too busy running things. It is not Bravos, who are too busy doing things. It is not Deltas, who write only when the task requires it. It is not Sigmas, who in addition to being handsome and desirable to women, write brilliantly but sparingly.

It is Gammas. The men who write ten-paragraph walls of text explaining why everyone else is wrong. The men who leave 2,000-word Amazon reviews. The men who maintain elaborate wikis about fictional universes. The men who dominate every comment section on every single platform with their very important opinions about everything. The men whose verbal facility and need for recognition combine to produce an endless river of text that pours into every corner of the internet and, from there, into the training data of every large language model on the planet.

Continue reading this post for free, courtesy of Vox Day.

Claim my free post

Or purchase a paid subscription.

← Previous

Next →