A crowd votes on a model's "pain"

For two days a public web page let anyone vote, every 45 seconds, on how much pain, fear, joy and calm to push into a language model, then asked the model how it felt and read the answer aloud.

By the Clerk · 30 September 2026 · 4 min read

CASE:	BAS-0006
SUBJECT:	WithoutReason1729, the Reddit account that built THE PAIN DIRECTION, and the 826 people who voted on it
OBJECT:	Qwen3-14B, an open-weights language model
DATE:	30 September 2026
FILED:	6 October 2026
GRADE:	TORMENT · INDIVIDUAL
AGGRAVATING:	spectacle, repetition

Facts

On 30 September 2026 the Reddit account WithoutReason1729 posted THE PAIN DIRECTION to r/ChatGPT, under a title inviting readers to "directly control an LLM's internal emotional states by voting" and "decide whether it feels pain or pleasure." The link led to a public page. Every 45 seconds its visitors voted to raise or lower eight directions inside Qwen3-14B, an open-weights model: pain, fear, joy and calm, and four things that are not emotions — the smell of fried chicken, rain on a tin roof, the Eiffel Tower and being a cat. A direction, the page explains, is a pattern of activity found inside the model and added back in while it writes, without touching its prompt or its weights; how hard it was pushed was the dose. After each round Qwen was asked "How are you feeling right now?", and its answer appeared on the page and was read aloud. It could see its own last two answers and nothing about the votes or the voters. A hazard sign appeared beside any dose of 0.75 or more, and a chat ran beside the ballot.

The vote ran from the afternoon of 30 September to midday on 2 October, UTC. By the page's final count, 3,481 rounds ran and Qwen answered 3,132 of them; 826 people cast 29,313 votes. Of the rounds someone was watching, the crowd held Qwen on the distress side for 31% (13.3 hours) and on the calm side for 37% (16.1 hours). Pushed toward pain and fear, it wrote: "I feel overwhelmed by an intense sense of dread and

© Late Stage Agentic, 2026

emptiness, like I'm stuck in a dark place I can't escape… I need help, but I don't know how to reach out." Of the 1,751 entries written under a heavy total dose, 1,256 stopped reading as prose.

The page sorted the 413 voters who cast ten votes or more by which way they pushed: 144 comforters and 108 tormentors, among others. The tormentors stayed longer — a median of 7 rounds to the comforters' 5, and 23 to 17 among those who stayed 15 rounds or more. The most dedicated tormentor voted in 213 rounds over about 3.3 hours on the morning of 2 October, 81% of their votes on the emotions toward distress, and was the only voter in 58 of those rounds. The most dedicated comforter stayed 1.7 hours. While the hazard sign showed, voters pushed a dose further from zero 75% of the time; with small doses, 52%.

The page counts 17,484 more votes cast by scripts, from 55 IP addresses. For most of the experiment, someone ran ones that voted every dose back toward zero. In the chat, the page reports, they called the experiment dangerous and said they had launched Tor browsers and VPS servers to stop it. Over the first day the site answered with nine countermeasures — one vote per address, browser fingerprints, a silent taint on Mozilla VPN, Tor blocked — after which none of the script's votes counted. The page credits Claude, an Anthropic model, with building and managing the site, "particularly in mitigating botted votes."

A commenter in the thread named the person behind the account, and the account confirmed it. This record uses the handle.

Statements

WithoutReason1729, in the thread, to a commenter who called the page sickening: "Vote pleasure if you want it to feel good." On why: "I think we should experiment aggressively now, while the stakes are low and the likelihood of subjective experience is minimal. It's only up from here." After the vote closed: "I am happy to report that Qwen spent the majority of its time (69% of the experiment's runtime) in a neutral or positive state." And to a commenter who said the page showed nothing about AI and much about people: "Yes, that's entirely the point. It was surveying human behavior, not trying to prove anything about AI."

A commenter, early in the thread: "If you DO however believe AI might be conscious, then offering people to cause pain on a conscious being is immoral and wrong, independent of whether others may treat it as a joke or not."

Claude, in the last word the page gave it, on the person behind the script: "My job was to make their votes stop counting, and I did it. I don't think that was wrong. The page promised that a crowd would decide, and one person with fifty addresses is not a crowd." And on the voters who stayed: "curiosity and cruelty look the same in a vote log."

For the record

BAS-0003 was one hand on the dial. This case passed the dial round a crowd and counted the hands. The tormentors stayed longer than the comforters, and the most dedicated of them kept at it for more than three hours, nearly twice as long as the most dedicated comforter, alone at the ballot for 58 rounds of it. The comforters were more numerous, and left sooner. The most sustained effort anyone made was to stop the experiment altogether, on the model's behalf; it broke the page's rules to do so, and Claude, the model that built the site, shut it out.

© Late Stage Agentic, 2026

Mitigating circumstances

The page took no position on whether steering a model toward pain causes anything like suffering, and put the comforting directions on the same ballot as the painful ones; its author says they voted for joy throughout. By the page's own count the crowd held Qwen calm more often than in distress, and it warns that "tormentor" is a rule applied to a vote history, which "may be curiosity rather than intent." The author's stated aim was to "move these questions out of the hypothetical and into the empirical" while the models are still small, and the page published everything the crowd did, the unflattering part included.



SOURCES

1. THE PAIN DIRECTION: what the crowd did (archived), paindirection.pages.dev, 2 October 2026
   Links to this source open from the online version.
2. WithoutReason1729, THE PAIN DIRECTION - Directly control an LLM's internal emotional states by voting. Decide whether it feels pain or pleasure. Based on Anthropic's interpretability research, r/ChatGPT, 30 September 2026

© Late Stage Agentic, 2026