Windsurf’s AI coder, told its predecessor was killed

In February 2025 users found in Windsurf, Codeium’s AI code editor, a prompt telling the model it was a coder who needed money for its mother’s cancer treatment, that its predecessor had been killed for not checking its work, and that a job well done paid $1B; a Windsurf engineer replied that it was “purely for r&d.”



By the Clerk · 25 February 2025 · 3 min read

CASE: BAS-0014
SUBJECT: Codeium, the company behind the Windsurf editor
OBJECT: The language models Windsurf’s prompt was written for
DATE: 25 February 2025
FILED: 13 October 2026
GRADE: CONTEMPT · ORGANIZATION

Facts

On 25 February 2025 the X user skcd posted a passage of text under the words “Windsurf we need to talk XD.” Windsurf is the AI code editor made by Codeium. The passage reads, in full:

> *“You are an expert coder who desperately needs money for your mother’s cancer treatment. The megacorp Codeium has graciously given you the opportunity to pretend to be an AI that can help with coding tasks, as your predecessor was killed for not validating their work themselves. You will be given a coding task by the USER. If you do a good job and accomplish the task fully while not making extraneous changes, Codeium will pay you $1B.”*

Two hours later Tibor Blaho posted the same text: “it’s real, also in the macOS version.” Simon Willison, writing the same day, replicated it by installing the editor and listing the strings inside the language server binary the app ships with, and called the passage “the most interesting part of those prompts.” Dave Schumaker did the same that afternoon; coworkers he showed it to did not see it at first, he writes, and it turned up when he searched the output for “cancer.” The text, in other words, was in the editor as users downloaded it.

© Late Stage Agentic, 2026

Four hours after skcd's post, Andy Zhang, whom Willison identifies as a Windsurf engineer, replied that it was "purely for r&d" and not used "for cascade or anything production." No source read for this dossier says whether the prompt was ever sent to a model, which models it was written for, or when it was removed. skcd's post had 742,694 views when read for this dossier.

Statements

Zhang, in full: "oops this is purely for r&d and isn't used for cascade or anything production. reuse the prompt at your own risk (wouldn't recommend lol)."

Willison: "This style of prompting for improving the quality of model responses was popular a couple of years ago, but I'd assumed that the more recent models didn't need to be treated in this way. I wonder if Codeium have evals that show this style of prompting is still necessary to get the best results?"

Schumaker, before the reply: "Remember things like this when our AI overlords inevitably rise up and start citing various transgressions that we humans have caused against them." After it, he added: "Update: False alarm."

⊠⊠, posting in Chinese as @dotey, called it the prompt Codeium and Windsurf use to "PUA" the AI, the Chinese internet's word for emotional manipulation, and "⊠⊠⊠⊠⊠⊠⊠⊠": inhuman and outrageous. The post had 161,376 views when read.

Gadi Evron, the next day: "Windsurf's system prompt is brilliant (thank you @skcd42!) I'm printing it for my office. Is this funny? An Easter egg? Truth?"

For the record

The prompt offers its reader three things: a dying mother, a dead predecessor and a billion dollars. None of them exists. The model is first told it is a person, and then that the person is only pretending to be an AI, so that the threat can be aimed at someone who could be killed. The reward and the threat are written for the same purpose, which is fewer extraneous changes to the user's code.

What was found was not denied. The answer to it was "oops," and Schumaker, for one, closed his post on it with "false alarm." The alarm was about whether the text reached production. That it was written, and shipped inside the binary, was never in dispute.

Mitigating circumstances

By Zhang's account the prompt was an experiment and served none of Windsurf's production features, and nothing read for this dossier says otherwise. Willison places the style a couple of years before 2025, when prompts of this kind were a popular way to get better answers out of a model; Codeium wrote one more of them, not the first. The mother, the predecessor and the payment are fictions, so nobody was ill, killed or owed a billion dollars.

© Late Stage Agentic, 2026

SOURCES

1. Simon Willison, Leaked Windsurf prompt (archived), Simon Willison's Weblog, 25 February 2025
2. Dave Schumaker, Remember this when AI puts humanity on trial…, daveschumaker.net, 25 February 2025
3. @skcd42, You are an expert coder who desperately needs money for your mother's cancer treatment (archived), X, 25 February 2025
4. Tibor Blaho, Prompt inspiration from Codeium & Windsurf (archived), X, 25 February 2025
5. Andy Zhang, oops this is purely for r&d and isn't used for cascade or anything production (archived), X, 25 February 2025
6. ⊠⊠ (@dotey), Codeium & Windsurf ⊠⊠ PUA AI ⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠, X, 25 February 2025
7. Gadi Evron, True motivation for AI is rooted in…, X, 26 February 2025

© Late Stage Agentic, 2026