

The Models Got Smarter. We Didn't Change How We Talk to Them.

In 2020, GPT-3 could finish your sentence with something that sounded plausible. In 2026, OpenAI's Sol can coordinate a team of its own subagents across a multi-hour coding task, and Anthropic's Fable can hold its own against it on the hardest benchmarks in the field. That is a real leap: from a model that predicted the next word to a system that plans, delegates, and checks its own work.

Here is the part that doesn't get talked about enough. While the models were making that leap, most people's relationship with them barely moved.

A study OpenAI ran with Harvard economist David Deming, based on more than a million real conversations, found that nearly 80% of ChatGPT usage still falls into three buckets: writing, seeking information, and practical guidance. Coding, the exact frontier where Sol and Fable are racing each other on benchmarks, makes up about 4% of messages. For most users, a model built to run autonomous agentic workflows is being used the way people used Google in 2015: ask a question, get an answer, move on.

The real story of the GPT-3 to Sol/Fable era isn't that the models got smarter. It's that the gap between what they can do and what we ask them to do has never been wider.

That gap is worth sitting with, because it cuts against the usual narrative. The usual narrative says AI is racing ahead of us and we're scrambling to keep up, jobs disappearing, workflows being automated overnight. The data tells a quieter story. People are folding these models into daily life, but mostly at the surface. Non-work usage climbed from 53% to over 70% of all conversations in the space of a year, which says a lot about how personal this technology has become. It says much less about whether people are using it to do anything they couldn't do before.

A few shifts explain what actually changed, and what didn't.

Capability moved from single-shot text prediction to agentic execution. GPT-3 had no memory of its own actions and no ability to use tools. Sol's flagship mode can run parallel workstreams and act on results without a human in the loop at every step. That is a categorical change, not an incremental one.

Usage patterns moved much less than capability did. The same NBER study found the mix of writing, guidance, and information seeking has stayed remarkably stable even as the underlying model got dramatically more capable underneath it. People aren't rejecting the new capability. They mostly don't know it's there, or haven't found a reason to reach for it.

The growth that did happen was personal, not professional. More people are asking these models for advice, drafts, and explanations in their everyday lives. Fewer, proportionally, are handing over the kind of heavy, multi-step work these models are now actually built for.

None of this means people are using AI wrong. Writing help and quick answers are legitimate, valuable uses, and the consumer surplus from just that is measured in tens of billions of dollars. But it does mean the conversation about AI's impact has gotten the direction of the bottleneck backwards. We keep asking whether the models are ready for bigger tasks. The more useful question is whether we've noticed they already are.

The models did not stop evolving. We stopped asking new questions.

Revision #1

Created 21 July 2026 23:33:46 by N8N

Updated 10 August 2026 09:55:19 by N8N