GPT-5: Impressive Progress, But Are We There Yet?

I Asked GPT-5 to Summarise My Meeting. It Made Up Solutions That Were Not Needed

GPT-5 launched with fanfare about solving AI's biggest problems: hallucinations, people-pleasing tendencies, and context confusion. The benchmarks look impressive, the marketing is smooth, and everyone's talking about the "revolutionary" improvements.

But here's what happened when I actually tried to use it for something important: it assertively invented issues that didn't exist in my meeting, then proposed elaborate solutions to fix them. Even worse? It agreed with my own misinterpretations instead of correcting me.

Don't get me wrong, GPT-5 is genuinely impressive and represents real progress. But if you're planning to trust it with anything critical, we need to talk about what these "solved" problems look like in practice.

The Good News: Five Key Improvements

1. No More Model Choosing Anxiety Remember scrolling through that confusing list of GPT variants, wondering whether you needed the "fast", the “mini”, or the "reasoning" model? GPT-5 eliminates this decision fatigue with an intelligent router that automatically selects the best model for your query. It's like having a smart assistant that knows whether you need a quick answer or deep analysis.

2. Smarter Information Lookup GPT-5 has enhanced browsing capabilities and improved training to reduce those confident-but-wrong statements we've all encountered. The model is better at looking up real-time information and has been specifically trained to make fewer factual errors.

3. Less People-Pleasing One of the subtler but important improvements addresses "sycophancy", the tendency for AI models to agree with you even when you're wrong. GPT-5 has been trained to politely disagree when necessary, separating being helpful from being a yes-person.

4. Nuanced Safety Responses Instead of the binary "I can help" or "I absolutely cannot help" responses, GPT-5 uses a more sophisticated approach that evaluates the actual output rather than just the intent. This means fewer unnecessary refusals while maintaining appropriate safety guardrails.

5. Million-Token Context Window Perhaps the most impressive technical achievement is GPT-5's ability to process up to 1 million tokens, enough for entire codebases or lengthy documents. But more importantly, it includes "Search-Augmented Attention" to focus on relevant parts rather than getting lost in the vastness.

The Reality Check: Progress, Not Perfection

While these improvements are genuinely impressive, my recent experience suggests we shouldn't declare victory just yet.

The Hallucination Challenge Persists Just last week, I used GPT-5 to extract and summarise key points from a client meeting recording. Not only did it occasionally misinterpret what was said, but it went further, proposing solutions to problems that didn't actually exist in the meeting. The model was confidently creative when I needed it to be conservatively accurate.

Sycophancy Still Sneaks Through During the same project, I found myself nodding along as the model confirmed my interpretation of certain discussion points. It was only when I double-checked the audio that I realised we were both wrong, the model had agreed with my misunderstanding rather than correcting it based on the actual content.

What This Means for Users

These experiences don't diminish GPT-5's achievements, but they highlight an important truth: we're witnessing evolutionary progress, not revolutionary perfection.

For Business Users:

- Always verify critical information, especially for client-facing work
- Use GPT-5 as a powerful starting point, not the final word
- The improvements are real, but human oversight remains essential

For AI Enthusiasts:

- Celebrate the progress while maintaining realistic expectations
- These improvements represent sophisticated engineering solutions to complex problems
- We're moving in the right direction, but the journey isn't over

The Bigger Picture

What's remarkable about GPT-5 isn't that it's perfect, it's that OpenAI has systematically identified and addressed specific weaknesses. The automatic model routing alone represents a significant user experience improvement, and the enhanced context handling opens entirely new use cases.

The fact that I can now process entire documents while having the model intelligently focus on relevant sections is genuinely transformative for my workflow, even if I still need to fact-check the outputs.

Looking Forward

As we integrate these more capable AI tools into our work and lives, the key is calibrating our expectations appropriately. GPT-5 represents meaningful progress on some of AI's hardest problems, but it's progress within a longer journey.

The models are becoming more reliable, more nuanced, and more capable, but they're not yet ready to work completely unsupervised. And perhaps that's exactly where they should be powerful enough to dramatically enhance our capabilities but still requiring the critical thinking and oversight that humans bring to the table.

What's your experience been with GPT-5? Have you noticed these improvements in your own use cases?

Note: I had already agreed with my counterpart that I was going to use ChatGPT to summarise the meeting and therefore no sensitive information was disclosed to OpenAI or its partners.

Disclaimer and Disclosure

Third-party Content and AI Assistance: This article references tools and software that are publicly available and proprietary to their respective creators. The author does not claim ownership or affiliation with these third-party products. This article was written by the author with assistance from Generative AI Language Models.
Transparency Notice: While every effort has been made to ensure accuracy, readers should verify information independently and consult official sources or documentation for the mentioned tools and software. The use of AI in the writing process is disclosed in the interest of transparency, but all opinions and analyses are the author’s own unless otherwise stated.