

Writing with “AI Partners”

The Perspective of an Aspiring Novelist

Landen Gonzalez



“Lilliput Lyrics,” Charles Robinson. British Library. Public Domain.

I have been a writer for the last six years and have worked on the same story continuously—at least weekly—since I was fourteen years old. After facing disinterest from friends and family and having a poor connection with other writers, once advertisements like those from Google (2025), OpenAI (2025), and Anthropic (2025) began to circulate lauding the potential of generative AI models to act as dedicated writing partners, I was excited to finally have a second set of ‘eyes’ on my work, so to speak. Unfortunately, it has become obvious over the past months that the ads overpromised on the abilities of chatbots. In what follows, I share my experiences for the benefit of other writers.

In all fairness, a writer aiming to use these chatbots as a stand-in for line-editors might find themselves quite satisfied. It is rare that a chatbot’s feedback regarding pacing,

syntax, or similar techniques is baseless or unhelpful. So long as one can cultivate a strong understanding of their strengths and limitations—including the wisdom of knowing when to reject a chatbot’s recommendations—generative AI models can prove invaluable to the writing process. However, as I am an English major with plenty of experience in writing and literature programs, this was not my main priority; I use *The Elements of Style* for reference when I’m in doubt. In my experience, chatbots are good at questions of technique because they are relatively objective, extensively documented, and answer to an agreed-upon set of expectations. To this extent, chatbots can regurgitate the rules for good writing as they do hard facts in other subjects.

But I was looking for a stand-in for beta readers. For example, I was interested in seeing if the themes of the work properly manifested across characters and their arcs; I wanted to know what would be assumed and understood about worldbuilding and characterizations. In short, I wanted to discuss the impression the chatbot had of the story. I wanted to ask, “So what do you make of all that? What do you think?” I wanted a reader.

To this end, I found the context window information for ChatGPT-5.2 (OpenAI, n.d.), Claude Sonnet 4.6 (Anthropic, 2026), and Gemini 3.1 Pro (Google, n.d.), in order to be sure that I chose a model designed to handle my 39,000-word (260-page) novel. Based on Google’s claim that 1 million tokens is equal to about 1,500 pages of text, I was confident that I was well within the limit and would find in the aforementioned chatbots great writing partners. And yet, even when I was working with a single chapter of no more than twenty pages, I experienced failures like that which I describe below. In other words, Google’s Pro chatbot seemed unable to “read” my work reliably *despite* the large context window.

When asked, Google’s Gemini reported that 3.1 Pro was the best model for my needs. I asked all three of the above chatbots the same question: “*This is the full novel, please tell me in exhaustive summary all plot threads and what each of the chapters is about. Take your time, be thorough.*” Gemini 3.1 Pro proceeded to misrepresent major plot beats, such as claiming that a principal character was killed in chapter six once they were shot, despite the fact that the character is still active in chapter seven—a crucial detail that Gemini omitted. Likewise, the main plot is driven by a prisoner’s pragmatic mindset towards survival; when he at one point discards a chance at armed rebellion, Gemini 3.1 Pro mistakenly represented him as having *hid* the weapons. That would

completely change the thematic and emotional stakes of the primary conflict, not to mention the entire characterization of the main protagonist. In addition, the reasoning and methodology of the story's climactic betrayal was confabulated even though other parts of the summary were accurate and most errors were omissions, not additions. To be fair, however, the rest of the mistakes were minor, and the summary preserved the vast majority of character work, worldbuilding, and plot progression.

At the time of these experiments (on March 9, 2026), Gemini performed the best. However, the problems I've described are significant. What, for a competent human synopsis, could be deliberate omission for the sake of punchiness, was in the case of Gemini's summary genuine gaps in understanding. Further probing revealed that Gemini repeated the same mistakes (about the shot character, for example) even when I had moved past asking for a summary. As far as I can tell, the incomplete summary either was or became the conversation's foundation. In addition, Gemini Canvas's AI was *far* more prone to "hallucinations"—the most embarrassing performance overall considering its explicit marketing as a creative partner. Most of the "information" it invented seems to have come from chapter titles alone.

Like Gemini, Anthropic's Claude 4.6 Sonnet began by proclaiming itself to be the best model for my purposes. The summary that it provided is absolutely riddled with minor instances of nonsense throughout, such as inferring that the description of a man's "yellowed eyes" and "wiry beard" referred to a black African with a "yellowed beard." The major issues were even more bizarre. For some reason, Claude 4.6 Sonnet could not discern the endings of each chapter: it would explicitly say that the end of each "cuts off." My best guess is that Claude was processing only the first two-thirds of each chapter and possibly confabulating the rest, which led to some truly baffling nonsense at the climax, when all these hallucinations converged. The most egregious mistake was its framing of the central betrayal as not a betrayal at all. According to Claude, it was a calculated trick actually serving the protagonists, a bold swing that it makes no attempt to reconcile with the other details of the plot that it got right. This fundamental misreading of the skeleton makes any more detailed, worthwhile discussion nearly impossible.

Finally, the worst chatbot, but probably the most entertaining and strange in its failure, was OpenAI's ChatGPT-5.2. In response to the same prompt as all the other chatbots, ChatGPT failed to serve up even one single element, character name, or reference from a document it had supposedly 'read'. Instead, it fabricated a vague structure and

summary that seemed to be based on archetypes from generic fantasy or hero's journey plots. The resemblance to the source material was so slight that I thought there might have been a corrupt PDF to blame, so I gave GPT-5.2 a second chance with a DOCX file instead. I also emphasized the problems I had noticed, including the lack of specificity and the substitution of generic fantasy tropes. The model's clever response was a second confabulation, only this time filled in with more fabricated character names and generic details. Again, there was no hint of anything remotely related to the story; my best guess is that the output responded solely to the title—*The Blessed & The Basic*.

Strangely, ChatGPT-5.2 was (during a prior trial) able to churn out an analysis about as full of holes as Claude's. The model will tell you the solution is to paste the full text (that is, an entire book) by hand into the chatbot. However, even on the monthly \$20 Plus plan, there is a limit to how much each message can hold and how many can be sent at a time. This means that you will surely hit said limit while simply trying to provide the chatbot with the full text.

Is this nitpicking? After all, two of the chatbots (and ChatGPT in its best result) accurately presented the general plot beats and basic ideas. The trouble is, no creative writer needs AI to regurgitate the broad outline of their work. They know that already, along with much more—they wrote it. For a writer like me, these chatbots ended up as only a waste of a paid subscription and the electricity and water necessary for generative AI models to function.

Indeed, what creative writers need most is specific and meaningful *engagement* with their work; that starts with a model that can accurately summarize the text. For example, if I were to ask a model about character psychology or the progression of an arc in hopes of getting useful feedback, a complete understanding of where someone was and how they reacted in a given scene could turn out to be the defining difference—something that human readers would understand as an important element in creating nuanced characterization. That may explain why chatbots, in my experience, become more nonsensical and 'hallucinatory' once they are prompted to deliver any deeper discussion.

So should writers in need of "AI Writing Partners" give up on this new technology, at least for now—with a possible exception for line edits? As I write, I believe that Gemini 3.1 Pro is the best of the worst for any meaningful help—but it is still plainly bad. Any

writer who approaches these systems in hopes of receiving useful feedback must allow for and expect the mistakes. Given that the chatbot is not actually ‘reading’ at all (but is instead making statistical predictions), any truly complete impression of the story is impossible. However, there might still be some use for the technology in creative writing. For example, do the model’s mistakes about the story reveal an interesting angle worth exploring instead? Even if chatbots are fundamentally incapable of human-like understanding by virtue of their design, it is possible that their misunderstandings and assumptions can offer some useful cues for clearer and more intentional writing. It is important to accept that “AI Writing Partners” are nothing like a beta reader and still less like an actual writing partner. But some writers might still find the feedback worthwhile despite the costs—both to their time and the environment.

Works Cited

- Google. (2025, March 18). *Introducing Canvas in Gemini*. Youtube.
http://youtube.com/watch?v=4Leardp_AGc
- Google. (n.d.). *Dive into large files and code repositories*. Gemini.
<https://gemini.google/overview/long-context/#:~:text=Dive%20into%20large%20files%20and,30K%20lines%20of%20code%20simultaneously>
- OpenAI. (2025, April 8). *Writing with canvas in ChatGPT*. Youtube.
<https://www.youtube.com/watch?v=Ks1C-5p04PY>
- OpenAI. (n.d.). *GPT-5.2 Model*. OpenAI Developers.
<https://developers.openai.com/api/docs/models/gpt-5.2>
- Anthropic. (2025, November 26). *Turning Claude into your thinking partner*. Youtube.
<https://www.youtube.com/watch?v=g3jc9SmkfMU>
- Anthropic. (2026, February 17). *Introducing Claude Sonnet 4.6*. Anthropic News.
<https://www.anthropic.com/news/claude-sonnet-4-6>