

Merlin for Red Teams

By Action Dan

Archived: 2026-04-02 12:44:57 UTC

Welcome back Internet people! Lately I've seen a rise in AI generated articles, blog posts, and even book content. I need to say loudly, as a reader, **this is a major turnoff**. If a reader can tell that something was written by AI, then the tools are being used poorly. **Please don't pass off LLM output directly as human writing**. It makes your work output feel cheap. AI should be used as a writing tool, it shouldn't be replacing human writers altogether.

When writers use LLM output verbatim the result is often stale and lacks clarity. In many cases, it actually makes ideas harder to understand. Current LLMs struggle to maintain consistent, logical models of complex ideas. So while the writing may sound polished at first, it sometimes misrepresents concepts or drifts into conflicting, multiple definitions. Moving past the coherency issues, it's often obvious when a writer has an overreliance on verbatim LLM output. There are many obvious *tells*. From the overuse of the em dash, to the nonsensical use of the colon; AI generated content sticks out to those who use frontier models often. Certain phrasing patterns also stand out. For example: "It's not X, it's Y." As a writer, this often feels like filler. Just write about Y. Just because these are the current form of these *tells* doesn't mean these are universal or ubiquitous *tells*. Quite the opposite, these will change over time as the models change, but the heavy users of the models will very likely recognize their output when used verbatim.

Don't get me wrong, *I'm not saying don't use LLMs to help you write*. I previously wrote about [how to use AI in your technical writing](#), such as creating templates, voice files, and dynamic prompts to generate rich content. It also makes for a great editor! But one of the key takeaways there is in the last paragraph, where I emphasize heavily modifying and adapting the output. You can't use the output verbatim; frontier LLM output is just too recognizable.

I recently read this great and thoughtful article titled "[Don't Let AI Write For You](#)", where Alex Woods lays out that the point of writing is to develop and cement thoughts worth communicating, not simply generating words or content. I couldn't agree with this more. I often use LLMs to help expand on ideas, or think about edge cases I might be considering. I use it to help me refine my writing prompts and generate starting points. But very rarely do I use the ideas or output verbatim. It's an incredibly useful tool, but in my opinion it shouldn't replace the art all together.

So I'll repeat it, and I hope somewhere out there other writers take it to heart. When writers use LLM output verbatim it comes across as incredibly lazy. And frankly, why would anyone read that? A reader could just prompt the model themselves and get the same result.



Source: <http://lockboxx.blogspot.com/2018/02/merlin-for-red-teams.html>