

# Underwriting AI: Is It Aligned With Insurance Industry's Best Interests?

Underwriters could find themselves attracted to an AI application that may appear useful. But that application could, in turn, prove to be damaging.

By **Robert Reville** and  
**David Loughran**

**T**he rapid advancement of artificial intelligence is eliciting a wide range of predictions for its impact, from solutions to our most pressing problems like cancer and climate change to the availability of omniscient personal assistants to widespread misinformation and the disempowerment of humanity. These predictions are on the minds of casualty underwriters as they review a growing number of submissions from businesses claiming to be powered by AI.

Perhaps no plaintiffs will be around to collect damages, but no one wants to have underwritten the AI application responsible for the disempowerment of humanity.

While the artificial general intelligence that could precipitate an AI apocalypse is likely years away, it's reasonable to be concerned about the harm that AI systems can cause today. Large language models (otherwise known as LLMs) like OpenAI's ChatGPT have no mental model of the world and yet their ability to mimic human thought processes through statistical



Reville Loughran

*Best's Review* contributor **Robert Reville** is a co-founder and chief executive officer of Praedicat Inc. *Best's Review* contributor **David Loughran** is a co-founder, senior vice president of product and chief economist for Praedicat Inc. They can be reached at [bestreviewcomment@ambest.com](mailto:bestreviewcomment@ambest.com).



inference is stunning. It's easy to be lulled into thinking you're interacting with a human when you chat with an LLM, and that fiction will only deepen as the audio and visual sophistication of LLMs advance.

The trouble with LLMs, of course, is that they have no innate understanding of right and wrong. Once prompted, an LLM will dispense whatever information is deemed optimal by the data that trained it, regardless of the consequences. While LLMs can illuminate and delight, they can also generate utter nonsense, perpetuate bias and promote toxic ideology. Even if big tech succeeds in installing robust guardrails on today's LLMs—and the AI safety community is skeptical—it seems inevitable that LLMs with no guardrails whatsoever will proliferate as training costs plummet. How is an underwriter to judge the potential for an AI application to come off the rails?

One solution is to evaluate the alignment of the



business model underlying a given AI application with the interest of its customers. Take, for example, chatbots deployed to assist licensed professionals like therapists. Chatbots, like humans, may occasionally dispense bad advice that hurts rather than helps. But the chatbot application in this case is still fundamentally aligned with the patient since the intent is to help. And the businesses that deploy these chatbots are well incentivized to monitor for systemic sources of malpractice before it causes untold harm. Harming patients, after all, is bad for business.

Contrast the AI therapist application with social media's deployment of algorithms designed to maximize advertising revenue through user engagement. At the heart of hundreds of personal injury lawsuits filed against social media platforms is the allegation that their algorithms are indifferent to the content delivered and the effect of that content

on the mental health of users. More generally, algorithms may be well aligned with the interest of the decision-maker but may not be with the object of that decision-making. Researchers in 2019, for example, discovered that a widely deployed algorithm in health care systems systematically recommended less-intensive treatment for nonwhite patients.

The Center for Humane Technology, an AI safety nonprofit, says generative AI of LLMs has now launched a "race to intimacy"—a ruthless competition to build LLMs that substitute for human companionship. An LLM that is truly aligned with your interests could make for a great companion. Sure, it might make mistakes, just like a friend would, but that can be forgiven. Underwriting such a benevolent business model sounds like a great opportunity. Just make sure your new best friend is free of ulterior motives. **BR**