

Toy-maker Mattel accused of planning “reckless” AI social experiment on kids



Most obviously, AI models are still prone to hallucination, Kaur noted. And while Mattel’s AI toys are “unlikely to cause physical harm,” toys giving “inappropriate or bizarre responses” could “be confusing or even unsettling for a child,” he said.

For parents, the emotional ties kids make with AI toys will also need to be monitored, especially since chatbot outputs can be unpredictable. Another LinkedIn user, Adam Dodge—founder of a digital safety company preventing cyber abuse, called EndTab—pointed to a lawsuit where a grieving mom alleged her son committed suicide after interacting with hyper-realistic chatbots.

Those bots encouraged self-harm and engaged her son in sexualized chats, and Dodge suggested that toy makers are similarly “wading into dangerous new waters with AI” that

could possibly “communicate dangerous, sexualized, and harmful responses that put kids at risk.”

“This was inevitable—but wow does it make me cringe,” Dodge wrote, noting that Mattel’s plan to announce its first product this year seems “fast.”

Dodge said that right now, Mattel and OpenAI are “saying the right things” by emphasizing safety, privacy, and security, but more transparency is needed before parents can rest assured that AI toys are safe.

AI is “unpredictable, sycophantic, and addictive,” Dodge warned. “I don’t want to be posting a year from now about how a Hot Wheels car encouraged self-harm or that children are in committed romantic relationships with their AI Barbies.”

Kaur agreed that it’s in Mattel’s best interest to give parents more information, since “public trust will be vital for widespread adoption.” He recommended that the toy maker submit to independent audits and provide parental controls to reassure parents, as well as clearly outline how data is used, where it’s stored, who has access to it, and what will happen if their kids’ data is breached.

For Mattel, a bigger legal threat forcing responsible design and appropriate content filtering may come from any unintentional copyright issues arising from using OpenAI models trained on a wide range of intellectual property. Hollywood studios recently sued one AI company for allowing users to generate images of their most popular characters and would likely be just as litigious defending against AI toys emulating their characters.

<https://arstechnica.com/tech-policy/2025/06/mattel-sparks-fear-that-planned-chatgpt-fueled-toys-will-warp-kids/>