When Chatbots Go Rogue

Ioana Gorecki

June 7, 2023

Last week, a mental-health chatbot used by the National Eating Disorder Association suddenly began giving diet advice to people seeking help for eating disorders. The rogue chatbot had apparently been developed as a closed system, but the software developer rolled out an AI component to the chatbot in 2022. NEDA claims it was not consulted about the update and did not authorize it. The organization has now taken the chatbot offline.

This incident demonstrates the potential dangers companies face when employing AI chatbots to provide customer service and address consumer needs.

Regulators and law enforcement agencies are taking note. In recent blog posts and reports, both the CFPB and FTC have cautioned companies about over-relying on chatbots and generative AI to provide customer service and resolve consumer concerns.

CFPB Spotlights the Use of Chatbots by Financial Institutions

On June 6, the CFPB released a new issue spotlight on the use of chatbots by banks and other financial institutions. The report notes that banks have increasingly moved from "simple, rule-based chatbots towards more sophisticated technologies such as large language models ("LLMs") and those marketed as 'artificial intelligence.'" While these chatbots are intended to simulate human-like responses, they can end up frustrating consumers' attempts to obtain answers and assistance with financial products or services. Some of the CFPB's listed concerns are:

- **Limited ability to solve complex problems, resulting in inadequate levels of customer assistance** (for example, difficulty understanding requests, requiring use particular phrases to trigger resolution, difficulty knowing when to connect with live agent). The CFPB argues this is particularly concerning in the context of financial services, where consumers' need for assistance could be "dire and urgent."
- **The potential for inaccurate, unreliable, or insufficient information.** In contexts where financial institutions are required to provide people with certain information that is legally required to be accurate, such lapses may also constitute law violations.
- **Security risks** associated with bad actors' use of fake impersonation chatbots to conduct phishing attacks at scale, as well as **privacy risks** both in securing customers' inputted data or in illegally collecting and using personal data for chatbot training purposes.

The CFPB notes that is actively monitoring the market to ensure financial institutions are using chatbots in a manner consistent with customer and legal obligations.

FTC Raises Concerns Regarding Chatbots and "Dark Patterns"

The FTC addressed the intersection of chatbots and "dark patterns" in a recent blog post. (As

explained in more detail here and here, “dark patterns” are sometimes defined as practices or formats that may manipulate or mislead consumers into tacking actions they would not otherwise take.) The Commission is worried that consumers may place too much trust in machines, and expect that they are getting accurate and neutral advice.

The agency cautioned companies that using chatbots to steer people into decisions that are not in their best interests, especially in areas such as finance, health, education, housing, and employment, is likely to be an unfair or deceptive act or practice under the FTC Act.

In addition, the FTC warned companies to ensure that native advertising present in chatbot responses is clearly identified, so that users are clearly aware of any commercial relationships present in listed results. The blog was very clear that “FTC staff is focusing intensely on how companies may choose to use AI technology…in ways that can have actual and substantial impact on consumers.”

Given the regulators’ avowed interest in this space, companies should take care that their use of chatbots comports with this most recent guidance.