DATA&
SOCIETY
ALGORITHMIC IMPACT METHODS LAB

Pilot 3 Case Report: Spiritual wellness company, Zenith chatbot

"*Our team was developing a mobile app from within the matrix of a private equity backed company. Many common sense guardrails on developing AI-forward technology, especially one targeting youth, were discussed internally and brushed aside under the unrelenting objective of imminent return-on-investment. Enlisting support from critical AI researchers meant that the reckless prioritization of business interests over ethical impact was defined and internally documented. It was empowering as leverage in internal discussions to have this undeniable, professional documentation of the consequences.*
Collaboration between critical AI researchers and corporate endeavors help detail the need for regulations, even when it cannot control harmful commercial interests." — Zenith Product Manager

1. Overview – *The partners, location, system, and proposed uses.*

AIMLab partnered with a commercial wellness company to assess Zenith, a spiritual wellness chatbot developed by a company owned by a private equity firm. Zenith was developed by a company that connects users to psychics, life coaches, and spiritual experts for one-on-one consultations. The company also controls several psychic advice platforms, where users can chat with occult practitioners at a set rate. The AIMLab team tested the chatbot with Gen Z students at a private university in San Francisco. Zenith was designed with multiple personas and proposed uses: to provide tarot readings, astrological insights, relationship advice, and support with stress management.

2. Background – *The context and motivation.*

The project emerged in the context of growing reliance on AI-based conversational agents to fill structural gaps in mental health care and spiritual guidance, especially in the wake of the pandemic, when digital intimacy and remote support became normalized. Chatbots like Zenith promise accessible, personalized wellness and divination experiences, appealing to young users already accustomed to astrology apps and digital encounters with spirituality. However, their deployment often occurs without sufficient testing or impact assessment, despite risks around privacy, misinformation, and dependency. AIMLab's motivation was to pilot participatory

methods of algorithmic impact assessment that center the concerns of impacted communities, exploring not only how users interact with such systems but also the broader ethical, social, and labor implications of automating spiritual and wellness support.

3. People consulted – *Who was involved in the assessment and how they were selected.*

The assessment engaged 28 undergraduate students from a private university in San Francisco, most of whom were media studies majors enrolled in a Digital Cultures class. This group was intentionally selected because they reflect the target user base for Zenith: Gen Z students who are both familiar with digital media and representative of diverse ethnic and cultural backgrounds.

People not consulted include younger teens, older adults, or working professionals. While this kept the study relevant, it excluded perspectives on some features of accessibility, oversight, and differing wellness needs that may be valuable for future research.

4. Algorithmic Impact Assessment method: *These methods were used to surface anticipated impacts.*

AIM Lab organized online workshops where participants were invited to test different chatbot personas. Participants directly engaged with the Zenith chatbot, experimenting with its various personas and generating prompts that reflected spiritual, therapeutic, and everyday concerns. These sessions relied on community red-teaming (that is, in which users act as experts to probe vulnerabilities) and group discussions, encouraging students to surface their hopes, concerns, and expectations related to system use. By grounding the evaluation in lived experience and speculative scenarios, the method captured both immediate user harms and broader ethical and social implications.

Both workshops began with an introduction to the Zenith application, its core functionalities, and the use of pre-defined synthetic personas developed to model different user interaction experiences. Workshop activities centered on prompt-based evaluation, where participants generated inputs to probe persona responses and assess system-level attributes such as user interface usability, response accuracy, and alignment behaviors. In the first workshop, a facilitator screen-shared the interface while participants collectively proposed prompts, effectively simulating collaborative red-teaming of the personas. The second workshop extended this protocol to include additional personas fine-tuned and deployed by the wellness company after the initial session. Participants were allowed to engage in distributed

prompt–response testing, either in smaller breakout groups or independently, to produce more granular, user-specific evaluations of model performance, safety guardrails, and potential failure modes.

5. Anticipated impacts + community questions and concerns. *Participants asked these critical questions about the technology. They also anticipated the following impacts from this technology.*

Participants raised significant concerns about Zenith's broader social and ethical impacts. They questioned the reliability of its personalization and memory systems, noting that generic or inconsistent responses in sensitive contexts could undermine trust and affect how users interpret guidance. The use of multiple AI personas introduced risks around authenticity, framing, and the shaping of emotional relationships with technology. Participants identified potential harms related to privacy, data governance, and surveillance, emphasizing uncertainty about how sensitive disclosures would be stored and used. They also flagged the risk of emotional dependency and the displacement of human support networks, as well as weak safeguards in high-stakes situations such as self-harm or abuse disclosures. Finally, they pointed to misinformation risks and inadequate user education about system capabilities, stressing the need for stronger guardrails, transparency, and ethical design to mitigate these impacts.

Participants also had a number of questions about the technical operations of the system, and how those details grounded their experience of its use:

- What are the ethical implications of a profit-driven, gamified AI system providing spiritual guidance?
- What are the risks of dependency and habit formation?
- Will using this chatbot make people more likely to turn to AI rather than turning to human connection, like friends, family, or qualified professionals?
- How is my personal data being stored, and who has access to it?
- What guarantees do I have that my private conversations won't be misused or sold?
- Can I trust guidance from a chatbot that gives tarot readings or career predictions without evidence, e.g. will I have a misplaced sense of optimism?
- What safeguards are in place if someone uses the chatbot in a moment of crisis, such as self-harm or abuse?
- Does the chatbot actually remember past conversations, or is it giving me generic advice each time?

Surprisingly, participants disclosed that they were often more comfortable sharing private information with the chatbot than with a stranger. Their prior experiences with commercial AI systems, such as ChatGPT, contributed to this sense of ease. However, participants also raised concerns about misinformation, inadequate guardrails for sensitive topics, data security, and dependency. Despite questions and anxieties about these risks, users remained open to the chatbot's potential as a spiritual wellness guide.

6. Outcome from system developer or deployer – *Describe how the system developer or deployer responded to the assessment.*

The system developer, represented by the product manager for Zenith, approached the assessment primarily as a form of user experience testing, seeking feedback on chatbot personas, their usability, and interface—rather than deeper ethical or societal impacts. While participants raised concerns about privacy, dependency, and misinformation, the company's focus remained on refining chatbot personalities and gamifying engagement, including plans to unlock new personas over time to encourage continued use. This highlighted a tension between AIMLab's critical, participatory orientation and the developer's commercial priorities, as the assessment surfaced ethical risks that were not central to the company's deployment strategy.

Other factors shaping the developer's response included the company's commercial incentives to quickly launch the product, which limited willingness to address structural concerns like privacy protections or user dependency. Management's emphasis on speed-to-market and data-driven engagement strategies, such as creating gamified features to extend use, overshadowed critical findings from the assessment. Additionally, AIMLab researchers operated as external partners without direct influence over design decisions, meaning their insights were treated as supplemental to internal UX priorities rather than central to product development.

7. Lessons for AIA practice – *What did we learn about what methods work (and do not work) to shape AIA practice?*

This case showed that community red-teaming and participatory workshops are effective for surfacing risks that technical audits often miss, such as emotional dependency, privacy anxieties, and the risks of relying on a digital system for spiritual guidance. Inviting impacted users as experts yielded actionable insights into how chatbots are actually integrated into daily life. At the same time, the project highlighted limits: when working with commercial partners under pressure to launch quickly, community-driven findings may be reframed as free UX testing rather than taken up as structural critiques. Engagements with impacted users did not address other potential downstream impacts, such as the labor implications for the on-demand

psychic advisors who are part of the parent company's suite of products and who might be concerned that their job functions will be taken over by spiritual wellness chatbots. Thus, while participatory AIA methods work well to capture lived experience and ethical concerns, they are less effective at shaping outcomes without mechanisms that compel companies to address those findings.