News and Perspectives

You Can't Launch This: Trust as Infrastructure in Digital Behavioral Health

Trevor van Mierlo, JMIR Correspondent

Abstract

Digital behavioral health interventions frequently fail to scale, even when evidence-based and technically and operationally sound. In this *News and Perspectives* article, researcher, digital behavioral health platform founder, and JMIR Correspondent Trevor van Mierlo concludes a four-part series examining why this occurs, reporting on the foundational role of trust.

Key Takeaways:

- Trust and engagement are not the same thing.
- As AI systems become more adaptive and inferential, trust may need to be treated as infrastructure rather than a feature.

Trevor van Mierlo, DBA, is the founder and scientific architect of the Evolution Health platform and acts as a consultant for academic and commercial digital mental health initiatives. The views expressed in this article are his own.

While preparing to launch a resilience course for—and co-developed with—displaced Ukrainians, our lived experience collaborators identified a major problem, as described in a previous article.

The issue wasn't technical. Content had been translated and culturally adapted by displaced Ukrainians. Security audits were complete. The infrastructure was stable.

The issue was a short explainer video with an AI-generated Ukrainian voice-over. It immediately reminded collaborators of Russian propaganda videos circulating online.

The problem was trust.

As AI becomes increasingly integrated into behavioral health, the question is no longer simply whether systems work. Trust may be equally important.

My interest in trust led to several conversations with researchers and practitioners across cybersecurity, informatics, behavioral health, and public digital infrastructure. What emerged was that trust can fail at multiple layers—social, technical, institutional, and behavioral—even in systems that are clinically validated and operationally sound.

Intervention Engagement ≠ Trust

Tonia San Nicolas-Rocca, PhD, is Faculty Associate Dean in the College of Information, Data and Society at San José State University, where her work focuses on cybersecurity, digital trust, and patient-centric health systems.

In discussing trust, San Nicolas-Rocca notes that there can be a power imbalance in digital health engagement. People may use health systems because, quite simply, they have to. Their access to care, services, or institutional support is often dependent on continued participation.

For example, someone may complete an assessment, log into a portal, or continue engaging with a platform or system they distrust. But people who distrust systems may also behave differently within them. They may minimize symptoms, avoid sensitive disclosures, disengage selectively, or provide information they believe is safer rather than accurate.

This creates an important problem for digital behavioral health. Engagement metrics may tell us whether individuals continue participating, but not whether the underlying behavioral data remain authentic, complete, or clinically meaningful.

Speaking with San Nicolas-Rocca caused me to reflect on some of my recent work. For years, my research has focused on using behavioral economics to influence user engagement through small, tailored changes in language and tone. In many ways, this is an attempt to solve digital health's original sin.

But are users aware of this? And if they are, does that change the way they behave inside the system?

Consent Is a Snapshot

Behavioral systems collect enormous amounts of longitudinal data. Increasingly, they also generate behavioral phenotypes: patterns of interaction that reveal how individuals engage with digital environments over time.

These patterns may reveal far more than traditional demographic or clinical variables alone. The more people interact with a system, the more the system learns about what keeps them engaged. Every click, login, uploaded image,

completed assessment, or abandoned exercise teaches the system something.

Benjamin Schooley, PhD, is a professor in the Department of Electrical and Computer Engineering at Brigham Young University. He argues that future systems should most likely require "computable governance." In simple terms, if systems continue evolving after deployment by learning from user behavior, disclosures and governance should also evolve.

Privacy policies and consent buttons are snapshots. Schooley's concerns align closely with contextual integrity: the idea that trust depends not simply on whether your data is secure, but whether information flows that use it remain appropriate. To be relevant and sustain trust, consent may also need to be dynamic.

Enforceable Trust

Kavya Pearlman, MS, is a cybersecurity architect and AI governance expert whose work focuses on artificial intelligence, digital trust, and emerging governance risks. Pearlman's concern is not simply what systems collect, but "what is inferred, how long those inferences persist, and whether governance remains enforceable over time."

In Pearlman's view, trust cannot rely solely on privacy policies, consent forms, or institutional assurances. Instead, governance must become transparent and verifiable. If an AI system generates a behavioral risk score, modifies treatment recommendations, or triggers an intervention, organizations should be able to reconstruct exactly what data informed that decision, which model was operating at the time, and whether the system functioned within approved governance boundaries. Most behavioral health platforms today cannot produce this evidence—they have usage logs and vendor dashboards, not forensic trails.

Pearlman's proposed mechanism draws from outside health care: cryptographically verifiable audit trails, modeled on aviation flight recorders and financial transaction logs. These produce tamper-evident records of what data was accessed, which model was operating, and what autonomy level governed each decision. Digital behavioral health, she notes, often relies on rough analytics where it should have forensic-grade traceability.

As Pearlman noted during our discussion, "If you can't explain and defend a decision after the fact, you shouldn't be automating it in behavioral health care."

In this context, static disclosures alone may not be sufficient to sustain trust, particularly when systems continue learning from users long after initial consent has been provided.

Trust Begins Before Governance

Valeriya Ionan, MBA, former Deputy Minister of Digital Transformation of Ukraine, helped build Ukraine's public digital infrastructure during periods of both pandemic response and full-scale war. Reflecting on these questions, she challenged one of the central assumptions underlying current discussions about AI governance and digital trust.

"We often assume trust is a feature technology creates," she wrote, "while I increasingly think trust is social first and tech second."

> *Before discussing trust in AI systems, we may need to ask a simpler question: what creates trust in the first place?*
>
> Valeriya Ionan, MBA

Her observation reframes much of the current conversation. Before people evaluate privacy protections, consent structures, or explainability frameworks, they must first believe the surrounding institutions and systems are worthy of trust.

This becomes especially important in environments shaped by misinformation, weak public communication, low digital literacy, or longstanding institutional distrust. In these contexts, trust problems may begin long before users even encounter the technology.

When Trust Becomes Infrastructure

Joe Grzywacz, PhD, Associate Dean for Research and organizer of the Health TechQuity initiative at San José State University, noted that equitable implementation depends not only on whether systems are available, but whether individuals believe the systems themselves are worthy of engagement and participation.

Grzywacz emphasized that trust inherently requires vulnerability. In digital behavioral health, individuals disclose intimate personal information while increasingly ceding interpretive authority to systems, platforms, and algorithms they often do not fully understand.

As behavioral systems become increasingly adaptive and predictive, this asymmetry between users and infrastructure may become a source of mistrust.

The broader implication may be that trust itself is becoming infrastructure.

Launching With the Wrong Voice

Trust does not automatically emerge from efficacy studies, privacy statements, or institutional reputation. People may distrust systems that are technically sound, clinically validated, and fully compliant.

Future behavioral systems may require trust to be actively maintained rather than simply assumed. Governance structures and transparency will remain essential. But they may prove insufficient if systems fail to establish legitimacy within the social and contextual environments in which individuals experience them.

Digital behavioral systems may ultimately succeed or fail depending on not only what they deliver, but whether people trust them. In the AI era, trust may no longer be a secondary feature of behavioral infrastructure. Instead, it may become the layer everything else is built on: engineered, maintained, and auditable, yet ultimately judged by the people asked to rely on it.

Keywords: trust; artificial intelligence; digital health; mental health services; consumer participation; patient-centered care; organizational policy; health governance

Please cite as:
van Mierlo T
You Can't Launch This: Trust as Infrastructure in Digital Behavioral Health
J Med Internet Res 2026;28:e104731
URL: https://www.jmir.org/2026/1/e104731
doi: 10.2196/104731