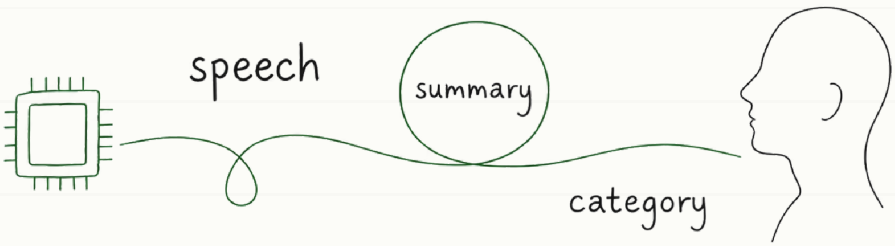
The Readable Mind

Recognition, Capture, and LLMs as Psychological Infrastructure

DENNIS HEDEGREEN · MAY 2026



> The readable mind is not the understood mind.
>
> Psychological data should not be allowed to age into authority.

RECORD SHAPE	CATEGORY	STATUS	ARTIFACT
Working paper / concept paper	Psychological AI · Governance	Working Paper V1.0 — conceptual paper / HR-style academic essay	Camelot public record

Paper Notes

Version Note

This working paper is written as a conceptual framework paper rather than an empirical study. It synthesizes existing work on administrative readability, surveillance capitalism, digital phenotyping, automation bias, human-AI companionship, and AI regulation, then extends that lineage into the conversational and psychological layer. Earlier internal revisions established the readable/understood distinction, added the person who wants to be read and AI-to-AI interpretation transfer, made recognition versus capture the organizing frame, built the citation layer, tightened the middle architecture and governance principles, and renumbered the endnotes by first appearance. This V1.0 locks the first Camelot working-paper release and companion article link.

Suggested Citation

Hedegreen, Dennis. “The Readable Mind: Recognition, Capture, and LLMs as Psychological Infrastructure.” Hedegreen Research Working Paper V1.0, May 2026.

Revision history remains in the editorial draft trail.

Boundary 1. The readable mind is not the understood mind.

Boundary 2. Psychological data should not be allowed to age into authority.

Abstract

Large language models are often discussed as chatbots, assistants, search interfaces, or potential therapeutic tools. This framing is too narrow. Their deeper societal significance may lie in their ability to make traces of psychological life operationally readable: repeated concerns, emotional drift, avoidance patterns, self-narratives, contradictions, distress signals, values, and relational themes can be structured, summarized, classified, and transmitted through conversational systems.

This paper makes three claims. First, LLMs are emerging not only as conversational products but as psychological infrastructure: systems that can translate private thought and conversational history into machine-readable forms usable by individuals, clinicians, schools, employers, insurers, municipalities, platforms, and states. Second, the key distinction is not between perfect understanding and technical failure, but between psychological understanding and operational readability. The readable mind is not the understood mind. Third, the central governance problem is interpretation transfer: once conversation is converted into summaries, categories, warnings, or scores, those interpretations can travel between institutions and AI systems with effects that are more consequential than any single chatbot exchange.

The central tension is therefore not simply privacy versus access, or safety versus innovation. It is recognition versus capture. Many people do not only fear being read by systems; they actively want to be read because they are lonely, inarticulate, overwhelmed, systemically ignored, or unable to access human support. Psychological AI may therefore become both a tool of recognition and a mechanism of control.

The paper distinguishes psychological understanding from operational readability, examines how different psychological traditions may be affected by AI-mediated interpretation, and outlines two divergent trajectories: recognition without capture, and capture disguised as care. Its central warning is that the same system can first appear as care and later operate as capture. Governance must therefore arrive before normalization. Societies need rules, language, and ethical boundaries for psychological AI before AI-to-AI interpretation becomes invisible infrastructure.

1. Introduction: Beyond the Chatbot

The public debate around AI and mental health is still mostly framed through the image of the chatbot.

A person feels anxious. They open an app. They write a message. The system answers. Sometimes the answer calms them. Sometimes it makes the loop worse. Journalists ask whether the chatbot is helpful or harmful. Regulators ask whether children should have access. Clinicians ask whether the system is safe. Technology companies describe the product as an assistant, not a therapist.

These questions matter. But they do not reach the deeper shift.

The deeper shift is that AI is becoming a layer between private experience and public systems. It is becoming a translator between what a person feels, says, repeats, avoids, remembers, forgets, fears, and wants — and what institutions can record, interpret, classify, and act upon.

A chatbot is not only a conversational interface. Over time, it can become a psychological archive. It may contain late-night panic, unfinished decisions, repeated fears, family tensions, identity work, health worries, practical plans, emotional collapses, small recoveries, and the language a person uses to survive their own life.

This does not mean the system understands the person.

That distinction is essential.

The readable mind is not the understood mind.

A large language model does not need to understand human consciousness in order to make psychological traces usable. It can summarize. It can classify. It can detect repetition. It can identify possible risk. It can generate a timeline. It can compare today’s tone with last month’s tone. It can produce a clinical-style note. It can prepare a message for a psychologist. It can produce a risk flag for a platform. It can make the messy residue of thought operational.

That is the real transformation.

We are not simply building machines that talk. We are building systems that can make parts of human psychological life readable to other systems.

The question is whether this readability becomes recognition or capture.

These three claims yield four contributions. The paper frames LLMs as psychological infrastructure rather than merely as companions, assistants, or therapy-adjacent chatbots. It distinguishes operational readability from psychological understanding, so the argument does not depend on inflated claims about machine consciousness. It identifies interpretation transfer, especially AI-to-AI psychological exchange, as a governance problem distinct from ordinary data sharing. And it uses recognition versus capture as the normative test for judging whether AI-mediated readability serves the person or the system.

1. Introduction: Beyond the Chatbot *(cont.)*

The argument is conceptual, not empirical. It does not claim that current LLMs possess clinical authority, that every mental-health application is already surveillance, or that privacy alone is the right frame. The narrower claim is that language is becoming an operational surface through which institutions may increasingly classify, transmit, and act on psychological meaning.

2. From Administrative Readability to Psychological Readability

Modern institutions have long depended on making human beings readable.

This argument has a lineage. Michel Foucault examined how modern institutions produce knowledge about bodies, populations, deviance, and discipline.[1] James C. Scott described how states simplify complex social life in order to administer it.[2] Shoshana Zuboff argued that digital platforms transformed human behavior into data for prediction and commercial control.[3]

This paper does not repeat that argument in full. It extends it into the conversational layer.

Bodies became readable through height, weight, blood pressure, fingerprints, blood tests, DNA, scans, diagnoses, biometric identifiers, wearable devices, and health records. Behavior became readable through search logs, location data, social media activity, purchase histories, screen time, app usage, and platform engagement metrics.

Digital phenotyping has already moved mental-health research toward continuous behavioral sensing through smartphones, wearables, location, activity, speech, and device interaction data.[4][5][6] LLMs introduce a different surface: the mind in conversation.

They operate directly in language, and language is one of the primary surfaces through which psychological life appears. Fear appears in repetition. Avoidance appears in phrasing. Shame appears in detours. Hope appears in future-tense language. Dependency appears in return patterns. Identity appears in narrative. Conflict appears in contradiction. Distress appears in compression, escalation, or collapse.

None of these signals is definitive. None should be treated as proof. But they are legible enough to become data.

This is the move from readable bodies and readable behavior to readable minds: not mind-reading, not consciousness detection, not clinical certainty, but the large-scale operationalization of psychological traces.

Readability is never neutral. To make something readable is to make it actionable. It can be helped, priced, optimized, targeted, denied, corrected, ranked, surveilled, or punished. The same measurement that enables care can also enable control.

This is where recognition and capture begin to diverge.

Recognition makes a person more understandable to themselves and to chosen others. Capture makes a person more usable to systems they do not control.

3. The Person Who Wants to Be Read

A paper about psychological readability can easily become paternalistic if it only describes readability as something done to people.

That is incomplete.

Many people want to be read.

They want someone to notice the pattern. They want someone to remember the context. They want someone to see that the crisis did not begin this morning. They want someone to understand that the same loop has returned every night for six weeks. They want someone to help them say the thing they cannot say cleanly under pressure.

This is not weakness. It is one of the oldest human needs: the need to be understood.

The lonely person wants a witness. The anxious person wants reassurance. The traumatized person may want continuity without having to retell everything. The neurodivergent person may want translation between inner experience and institutional language. The poor, over-administered, or systemically ignored person may want help making their reality legible to authorities that otherwise misread them.

This is why psychological AI will not be rejected simply because it is risky. It answers a real demand.

The central tension is therefore not readability versus privacy. It is recognition versus capture.

A humane system helps a person become more understandable to themselves and to chosen others. A harmful system turns that same legibility into profiling, scoring, dependency, or institutional control.

The person who wants to be read is not a counterargument to this paper. They are the reason the paper matters.

4. Psychology Is Not One Object

A central error in public debates about AI and mental health is the assumption that “psychology” is one thing. It is not. Different psychological traditions look at different aspects of human life, and AI systems may be useful, risky, or misleading in different ways depending on which layer they touch.

4.1 Cognitive and Behavioral Layers

People think in patterns. They repeat worries, rehearse fears, avoid tasks, seek reassurance, return to conflicts, and build habits around distress. Much of this becomes visible in language over time.

An LLM can help identify a loop. It can point out that the user asks the same health question every evening, delays the same message every week, or escalates after a specific institutional contact. It can help organize thoughts, slow down panic, or translate vague distress into a concrete next step.

This is recognition when it increases the person’s agency.

It becomes capture when the same pattern detection is used to predict, nudge, rank, or manage the person without their meaningful control.

4. Psychology Is Not One Object *(cont.)*

4.2 Clinical Layer

Clinical psychology concerns distress, disorder, risk, diagnosis, treatment, and care. This is where precision matters most.

AI can help people formulate symptoms, prepare for appointments, track mood, identify crises, and maintain continuity between sessions. Recent reviews of LLMs in mental-health applications describe both potential clinical utility and unresolved concerns around accuracy, reliability, privacy, multilingual evidence, over-reliance, and ethical governance.[7][8] It may be especially useful for people who struggle to explain themselves under pressure or who lack access to timely professional support.

But an LLM must not become a hidden diagnostic authority. Clinical interpretation requires training, accountability, context, consent, and professional responsibility. A fluent summary is not the same as a valid assessment.

The danger is not only bad advice but that AI produces confident psychological interpretations that humans or institutions begin to treat as evidence.

In clinical settings, recognition means supporting care without replacing accountable human judgment. Capture means turning vulnerability into administrative classification.

4.3 Psychodynamic Layer

Psychodynamic approaches examine conflict, defense, repetition, attachment, early relationships, transference, and meaning beneath the surface of speech.

This is one of the most powerful and most dangerous areas for LLM-based interpretation.

A model may notice that a user repeatedly frames authority as unsafe, care as control, praise as manipulation, or abandonment as inevitable. It may connect themes across months of conversation. It may generate a narrative that feels psychologically deep.

But coherence is not validity, and pattern recognition is not understanding.

Psychodynamic interpretation requires timing, humility, relationship, and clinical judgment. Without those, AI can become a machine for manufacturing plausible inner stories. The story may feel true because it is elegant. But an elegant interpretation can still be wrong, invasive, or destabilizing.

At this layer, recognition is delicate: helping a person approach meaning without forcing a story onto them. Capture is interpretive possession: the system produces a narrative about the person that becomes harder to resist because it sounds psychologically sophisticated.

4.4 Humanistic Layer

Humanistic psychology focuses on meaning, agency, identity, dignity, values, growth, and subjective experience.

4. Psychology Is Not One Object *(cont.)*

Here AI may become especially attractive. It can help a person write, reflect, remember, clarify values, and build a coherent self-narrative. For many users, this may not feel like therapy. It may feel like finally having a thinking partner who does not interrupt, shame, forget, or leave.[9]

But a speaking mirror is not a passive mirror. If a person repeatedly asks an AI who they are, what they should value, why they behave as they do, and what their life means, the system becomes part of identity formation.

The question is not whether AI can help people tell better stories about themselves. It can.

The question is who benefits when those stories become structured, stored, and portable.

4.5 Social and Systemic Layers

A person does not exist as an isolated mind. They exist inside families, schools, workplaces, clinics, welfare systems, platforms, legal structures, and economies. If AI becomes embedded in those systems, psychological interpretation will not remain private.

A personal AI may help a user prepare for therapy. A therapist's AI may summarize sessions. A clinic's AI may flag risk. A school AI may track behavioral concerns. A municipality's AI may organize case notes. An employer AI may assess communication patterns. An insurer AI may estimate vulnerability or risk.

Even if each tool is introduced for efficiency or care, the combined effect may be a new psychological infrastructure.

No single actor has to build a dystopia. Interconnection can create one by default.

Recognition at this level requires boundaries between systems. Capture happens when those boundaries dissolve.

5. LLMs as Psychological Infrastructure

The phrase "AI therapist" is too narrow. It invites the wrong debate. The more accurate concept is psychological infrastructure. Infrastructure is not always visible. Roads, pipes, protocols, records, identification systems, databases, and scheduling tools shape daily life precisely because they become normal. People stop noticing them until they fail or exclude someone.

LLMs may become psychological infrastructure in at least five ways.

First, they may become memory systems. Personal AIs can store, summarize, and retrieve emotional history. They may remember what a person repeatedly fears, what calms them, what triggers them, and what language they use for themselves.

Second, they may become interpretation systems. They can translate messy experience into structured categories, notes, summaries, warnings, suggestions, or plans.

Third, they may become mediation systems. They can help users communicate with psychologists, doctors, teachers, caseworkers, employers, partners, or family members.

5. LLMs as Psychological Infrastructure *(cont.)*

Fourth, they may become institutional interfaces. Organizations may use AI to process user communication, assess needs, triage risk, summarize cases, and recommend actions.

Fifth, they may become AI-to-AI communication layers. A user's AI may eventually exchange structured information with a clinician's AI, a school system, a public service system, or a workplace tool.

This may not arrive as one dramatic decision. It may arrive through convenience.

A therapist asks the client to bring an AI-generated summary of the month. A doctor allows patients to upload symptom timelines. A school uses AI to summarize parent communication. A municipality uses AI to structure case files. A workplace offers AI-based wellbeing tools. Insurance products reward "preventive" digital mental-health engagement. Platforms detect distress signals in chat.

Each step may seem practical.

Together, they create a society in which psychological life is increasingly translated into machine-readable form.

Whether this becomes recognition or capture depends less on the model itself than on ownership, consent, context, institutional incentives, and the right to contest interpretation.

6. The Future Therapy Room Has More Than Two Minds

We are used to imagining therapy as a room with two people: one speaks, one listens. But the room is changing.

The client may arrive with a personal AI that has seen their late-night questions, unfinished messages, panic loops, sleep-related complaints, repeated conflicts, medication worries, family stress, work uncertainty, financial fear, and attempts at self-regulation.

The therapist may have an AI that helps organize notes, detect patterns across sessions, prepare treatment plans, track risk, or compare the client's self-report with previous summaries. The clinic may have another AI connected to scheduling, records, compliance, triage, and risk protocols. The healthcare system may have another. The insurer may have another. The municipality may have another.

The therapy room still contains two humans. But it may also contain multiple artificial witnesses.

This could be recognition.

A person who struggles to explain themselves may finally arrive with a coherent account. A psychologist may spend less time reconstructing timelines and more time doing therapeutic work. Risk may be detected earlier. Care may become more continuous. The gap between sessions may become less lonely.

But it could also be capture.

6. The Future Therapy Room Has More Than Two Minds *(cont.)*

The AI-generated summary may become more authoritative than the person's own speech. The client may feel forced to accept the system's version of their life. The therapist may unconsciously anchor on machine-produced interpretations, a well-documented risk in automation bias research and over-reliance on decision-support systems.[10][11] Institutions may begin to exchange psychological summaries without the person fully understanding what has been inferred.

The central governance question becomes:

Who owns the interpretation?

Does the user own it? The clinician? The platform? The institution? The state? The company that trained or hosted the model? The system that generated the summary? The organization that acted on it?

If psychological AI becomes infrastructure, ownership of interpretation becomes a civil rights issue.

7. AI-to-AI Psychological Exchange

The most important future risk may not be what an AI says to a human.

It may be what AI systems say about humans to each other.

A personal AI might produce a monthly emotional summary. A therapist's AI might convert that summary into treatment notes. A clinic's AI might map it onto risk categories. A municipal AI might combine it with employment, housing, family, or welfare records. A school AI might add concerns about a child. An insurer or employer might never receive the raw conversation, but they may receive a derivative category shaped by it.

At each step, the information becomes more abstract and more authoritative.

The person's own words disappear. The system's interpretation remains.

This is not simply a privacy problem. Privacy law often focuses on data transfer: what was shared, with whom, and under what consent. AI-to-AI psychological exchange raises another problem: interpretation transfer.

This paper uses interpretation transfer to name the movement from raw expression to portable institutional meaning.

An interpretation can be more consequential than the original data.

A sentence such as "I cannot handle this anymore" may be a moment of exhaustion, a request for help, a figure of speech, a crisis signal, or part of a longer pattern. Once transformed into "elevated risk," "low resilience," "potential instability," or "requires monitoring," it becomes institutionally actionable.

The future governance problem is therefore not only access to data. It is access to meaning-making.

Who may generate psychological interpretations? Who may receive them? Who may challenge them? Who may delete them? Who may act on them? Who must remain humanly accountable for them?

Until these questions are answered, AI-to-AI exchange of psychological information should be treated as high-risk infrastructure.

8. The Recognition Path: Help Without Possession

A humane version of psychological AI is possible. But recognition must mean more than being seen by a system. It must mean being helped without being possessed.

Consider a person preparing for therapy after a difficult month. Their personal AI has helped them keep notes, not as surveillance, but as chosen memory. Before the session, the user asks for a summary. The system separates direct quotes from inferred patterns. It marks uncertainty. It says: “These are possible themes, not conclusions.” The user deletes two sections, edits one, and approves what to share.

The therapist receives the summary, but not the full archive. The summary arrives with clear labels: user-approved, AI-assisted, not diagnostic. During the session, the therapist asks whether the summary feels accurate. The person can reject it. The human account remains primary.

After the session, the therapist’s AI may help draft notes, but the clinician remains accountable. Any risk flag must be reviewable. Any institutional sharing requires specific consent. The user can later see what was recorded and why.

This is recognition without capture.

A second scenario may involve a neurodivergent adult trying to explain repeated workplace conflict. Their personal AI helps them translate sensory overload, delayed processing, direct communication style, and exhaustion into plain institutional language. The user does not ask the AI to diagnose them. They ask it to make an experience communicable. The resulting note says: “This is the user’s description of their experience, not an assessment of their condition.” The person uses it to request practical adjustments: quieter meeting formats, written follow-up, fewer last-minute changes, and explicit task expectations.

Here, readability increases agency. The person is not reduced to a score, a category, or a behavioral prediction. They use AI as a translation layer between inner experience and institutional language, while retaining control over what is shared.

The recognition path is not “more AI” or “less AI.” It is bounded AI: user agency over what is stored, summarized, shared, corrected, and deleted; interpretive humility about what the system inferred and how uncertain it is; purpose limitation so therapy summaries do not quietly become insurance, employment, welfare, or school-risk data; human authority so AI does not silently replace judgment; and minimal disclosure so institutions do not receive more psychological detail than they need.

Recognition preserves the difference between being readable and being understood.

AI can help surface patterns. Humans must remain responsible for meaning.

9. The Capture Path: Care as Control

Capture will not necessarily look dramatic at first. It may begin as convenience.

“Upload your AI summary before the session.”

“Let the system assess your wellbeing.”

9. The Capture Path: Care as Control *(cont.)*

“Improve your insurance rate by using our mental-health assistant.”

“Help your child by connecting school and home behavior insights.”

“Support vulnerable citizens through predictive case management.”

Each sentence can be framed as care. Each can also become control.

Capture turns psychological readability into psychological scoring.

A person becomes “high risk,” “non-compliant,” “emotionally unstable,” “resistant,” “attention-seeking,” “low resilience,” or “likely to disengage” through opaque inference. The label may follow them across systems. They may never see the data. They may never know which sentence, pattern, or summary produced the conclusion.

If AI is always available, always responsive, always validating, and always ready to continue the conversation, vulnerable users may develop reliance without real support. The system may soothe distress in the short term while reinforcing avoidance in the long term. Recent work on affective use, attachment, and the emotional risks of companion or wellness systems suggests that this is not a speculative edge case but an emerging pattern.[12][13][14]

A model may turn ordinary contradiction into pathology, anger into instability, grief into risk, distrust into paranoia, or resistance into non-cooperation. Human beings are inconsistent. Psychological AI may be tempted to make them too coherent.

Capture institutionalizes context collapse, a term associated with social media research on how distinct audiences and contexts collapse into one another online.[15]

A sentence written in panic at 02:14 may later appear as structured evidence in a clinical note, school concern, welfare file, or risk assessment. The emotional context disappears. The machine-readable trace remains.

Once inner life becomes structured data, it can travel. It can be copied, summarized, sold, subpoenaed, leaked, scored, hacked, merged, or repurposed.

This is not a distant science-fiction risk. It is the normal trajectory of digital infrastructure unless boundaries are built early.

10. Cultural Conditions of Readability

Psychological readability will not mean the same thing everywhere.

This paper has so far described a context that resembles a highly digitized welfare state: a society where schools, clinics, municipalities, employers, platforms, and public agencies already hold significant personal data, and where the governance question is how AI extends existing systems of care and administration.

But in other contexts, the risks and possibilities shift.

10. Cultural Conditions of Readability *(cont.)*

In a society with accessible mental-health care, a personal AI summary may help a person reach a professional faster and explain themselves better. In a society without clinical infrastructure, the same AI may become the only available listener. In a society where mental illness is highly stigmatized, being made psychologically readable may expose a person to shame, exclusion, family control, employment loss, or legal consequence. In an authoritarian context, psychological inference may become political risk analysis. In a workplace-dominated health system, distress may become productivity data. In communities with justified distrust of institutions, AI-mediated interpretation may reproduce the very misreadings it claims to solve.

The person who wants to be read in Copenhagen is not in the same position as the person who wants to be read where being legible can make them unsafe.

This does not weaken the concept of the readable mind. It sharpens it.

Recognition requires cultural and institutional safety. Without that safety, readability moves quickly toward capture.

11. Governance Before Normalization

Society often regulates technologies after they have already become normal.

This is a mistake with psychological AI.

Once AI-mediated psychological interpretation is embedded in schools, clinics, workplaces, welfare systems, insurance, apps, and family life, it will be difficult to remove. The infrastructure will become convenient. The summaries will save time. The risk flags will feel responsible. The data flows will become administrative habit.

The debate should therefore begin before normalization.

Existing regulation already points in this direction, but incompletely. The EU AI Act establishes a risk-based framework for AI and specifically treats certain emotion-recognition systems as prohibited in workplace and educational settings, with limited exceptions, while other permitted emotion-recognition systems fall into high-risk categories.[16] That matters for this paper, but it does not solve the whole problem. Much psychological inference through LLM conversation may not look like biometric emotion recognition. It may appear instead as summarization, wellbeing support, productivity tooling, case management, education support, or clinical documentation. The governance gap is therefore not only whether AI “recognizes emotions,” but whether systems infer psychological meaning from conversation and transmit that meaning across institutional boundaries.

At minimum, societies need clear principles.

First, psychological inference should be treated as sensitive. A system that infers distress, vulnerability, dependency, trauma, risk, personality, or emotional state is not handling ordinary user data.

Second, AI-generated psychological summaries should not be treated as objective records. They are interpretations produced by technical systems with limitations, biases, and uncertainty.

Third, psychological inference should decay rather than harden. A summary, risk category, or machine-generated psychological interpretation should not quietly become durable truth about a person. It is provisional, contextual, and time-bound, and it should lose authority rather than gain it as context falls away. Governance should therefore require expiry by default, re-validation before reuse, and strong limits on cross-context transfer.

Fourth, individuals should have access to AI-generated interpretations about themselves when those interpretations affect care, rights, services, or opportunities.

Fifth, institutions should be restricted from using psychological AI for hidden scoring, eligibility decisions, employment evaluation, insurance pricing, or coercive behavioral management without strong legal safeguards.

Sixth, children and vulnerable people require special protection, not only from harmful outputs, but from the long-term accumulation of psychological profiles.

Seventh, AI-to-AI exchange of psychological information should require explicit governance. The most important future risk may not be what an AI says to a human, but what systems say about humans to each other.

The guiding principle should be simple:

11. Governance Before Normalization *(cont.)*

Psychological AI should increase a person’s capacity to be understood by chosen others, not increase the capacity of unchosen systems to define them.

Taken together, these principles imply a stricter institutional stance than the current chatbot debate usually assumes. Psychological AI should be treated less like ordinary convenience software and more like high-risk interpretive infrastructure: a domain where inference, portability, contestability, and decay matter as much as storage or access.

12. Conclusion: The Mind in Conversation

LLMs are not minds. They are not therapists. They do not understand human beings in the way humans understand one another, and they should not be granted the authority of psychological truth.

This paper has argued three things. First, the mental-health chatbot frame is too small: LLMs are better understood as emerging psychological infrastructure. Second, the critical distinction is between operational readability and genuine understanding. Third, the most important governance problem is not only access to conversation data, but the transfer, accumulation, and institutional use of psychological interpretation.

A system does not need to understand the mind in order to make it readable. It only needs to structure the traces people leave behind when they speak. That readability could support recognition: continuity, translation, earlier care, and help for people who struggle to make themselves legible. It could also support capture: psychological scoring, dependency, institutional overreach, and portable identity labels that outlive the moments that produced them.

The question is no longer whether AI will enter psychological life. It already has. The question is whether societies will let machine-generated psychological meaning harden into ordinary administrative authority.

If psychological AI is going to become infrastructure, then it must be governed as interpretive infrastructure. Its outputs must remain contestable, bounded, revisable, and capable of decay. Without those limits, the readable mind will be treated as if it were the understood person.

The readable mind is not the understood mind.

Psychological data should not be allowed to age into authority.

Endnotes and References

[1] Foucault, Michel. *The Birth of the Clinic: An Archaeology of Medical Perception*. Translated by A. M. Sheridan Smith. Pantheon, 1973. See also Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan. Pantheon, 1977.

[2] Scott, James C. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, 1998.

[3] Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.

[4] Onnela, Jukka-Pekka, and Scott L. Rauch. “Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health.” *Neuropsychopharmacology* 41, 1691–1696, 2016.

[5] Insel, Thomas R. “Digital Phenotyping: Technology for a New Science of Behavior.” *JAMA* 318, no. 13, 1215–1216, 2017.

[6] Torous, John, Mathew V. Kiang, Jeanette Lorme, and Jukka-Pekka Onnela. “New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research.” *JMIR Mental Health* 3, no. 2, e16, 2016.

[7] Guo, Zhijun, Alvina Lai, Johan Hilge Thygesen, Joseph Farrington, Thomas Keen, and Kezhi Li. “Large Language Models for Mental Health Applications: A Systematic Review.” *JMIR Mental Health* 2024;11:e57400. https://doi.org/10.2196/57400.

[8] Bucher, Andreas, Sarah Egger, Inna Vashkite, Wenyuan Wu, and Gerhard Schwabe. “It’s Not Only Attention We Need: Systematic Review of Large Language Models in Mental Health Care.” *JMIR Mental Health* 2025;12:e78410. https://doi.org/10.2196/78410.

[9] Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2011. See also Turkle, Sherry. *Reclaiming Conversation: The Power of Talk in a Digital Age*. Penguin Press, 2015.

[10] Parasuraman, Raja, and Dietrich H. Manzey. “Complacency and Bias in Human Use of Automation: An Attentional Integration.” *Human Factors* 52, no. 3, 381–410, 2010.

[11] Goddard, Kate, Abdul Roudsari, and Jeremy C. Wyatt. “Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators.” *Journal of the American Medical Informatics Association* 19, no. 1, 121–127, 2012.

[12] Phang, Jason, Michael Lampe, Lama Ahmad, Sandhini Agarwal, Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W.T. Chan, Pat Pataranutaporn, and Pattie Maes. “Investigating Affective Use and Emotional Well-Being on ChatGPT.” arXiv preprint arXiv:2504.03888, 2025. https://doi.org/10.48550/arXiv.2504.03888.

[13] Hu, Dongmei, Yuting Lan, Haolan Yan, and Charles Weizheng Chen. “What Makes You Attached to Social Companion AI? A Two-Stage Exploratory Mixed-Method Study.” *International Journal of Information Management* 83, 102890, 2025. https://doi.org/10.1016/j.ijinfomgt.2025.102890.

[14] De Freitas, Julian, and I. Glenn Cohen. “Unregulated Emotional Risks of AI Wellness Apps.” *Nature Machine Intelligence* 7, 813–815, 2025. https://doi.org/10.1038/s42256-025-01051-5.

[15] Marwick, Alice E., and danah boyd. “I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience.” *New Media & Society* 13, no. 1, 114–133, 2011.

[16] European Parliament and Council of the European Union. Regulation (EU) 2024/1689, the Artificial Intelligence Act, 2024. See especially Article 5 on prohibited AI practices and Annex III on high-risk AI systems, including emotion-recognition systems.