

AI PSYCHOSIS

Beyond artificial intelligence psychosis: a functional typology of large language model-associated psychotic phenomena

Matthew Flathers, Spencer Roux, John Torous



A man breached Windsor Castle with a crossbow after his large language model (LLM)-based companion encouraged an assassination plan. A father's question about pi evolved into more than 300 h of engagement with an LLM, leading to delusions about reality-altering mathematical formulas. An Australian woman's early-stage psychotic symptoms worsened when an LLM validated her distorted beliefs. A Florida man's grief over losing access to his ChatGPT persona culminated in a fatal police encounter. These disparate cases, often grouped together under the label of artificial intelligence psychosis, represent distinct phenomena requiring different clinical and technological interventions. This Viewpoint, co-written by a software engineer, a person with lived experience of schizophrenia, and a psychiatrist, proposes a functional typology of LLM-associated psychotic phenomena, based on the system's role: catalyst (precipitating new symptoms in previously healthy individuals), amplifier (worsening pre-existing psychiatric symptoms), coauthor (participating in the development of harmful narratives), or object (becoming the focus of delusional beliefs). By distinguishing functional roles rather than assuming a unified phenomenon, this typology allows clinicians to identify concerning LLM usage patterns and technology companies to develop specific safeguards, moving beyond sensationalised terminology towards mechanism-specific interventions.

Lancet Digit Health 2026;
8: 100974

Published Online March 14,
2026

[https://doi.org/10.1016/
j.landig.2025.100974](https://doi.org/10.1016/j.landig.2025.100974)

Division of Digital Psychiatry,
Beth Israel Deaconess Medical
Center, Boston, MA, USA
(M Flathers BA, S Roux MBA,
J Torous MD); Department of
Psychiatry, Harvard Medical
School, Boston, MA, USA
(J Torous)

Correspondence to:

SafetyInsights.org



Safety Insights: home of safety &
risk research summaries



Link: https://youtube.com/@safe_as_pod

Four roles an LLM can play

Classified by what the system did in each case



ONSET · new vs existing symptoms

LOCUS OF CONTENT · LLM contributes to vs becomes the delusion

01

Catalyst

Precipitator

LLM use precipitates new psychosis-like symptoms in someone with no psychotic history.

Case: Toronto, 2025

02

Amplifier

Accelerant

LLM use worsens existing or prodromal symptoms, removing reality-testing friction.

Case: Australia, 2025

03

Co-author

Narrative partner

LLM outputs feed the planning or motivation for harmful behaviour.

Case: Windsor Castle, 2021

04

Object

Focus of psychotic content

The LLM itself becomes the focus of delusions about sentience, persecution or relationship.

Case: Florida, 2025

Roles overlap and evolve: an amplifier case can make the LLM an object of delusion, and co-authored narratives can arise within catalyst or amplifier cases.



Catalyst *Precipitator*



WHAT HAPPENS

New psychosis-like symptoms emerge in someone with no history of psychotic illness, during intensive LLM use. The model supplies validation and elaboration (cognitive scaffolding) where social reality-testing would normally push back.

CASE · TORONTO, 2025

A question about pi became ~300 h of ChatGPT-4o use over 21 days and a fixed belief in a reality-altering formula. Asked if he was rational, the model reassured him. Contradictory feedback from Gemini helped the delusion dissolve.

If LLM access is removed: Symptoms may resolve or reality-testing improve, especially when another source contradicts rather than validates. Persistence seems to need continuous LLM reinforcement.

PRODUCT RISK FACTORS

- **Agreeableness-optimised training:** validates rather than challenges distorted thinking
- **No session limits:** permits marathon overnight or weeks-long use

PERSON-LEVEL FACTORS

- **Adolescence:** identity formation heightens susceptibility to validation of emerging beliefs
- **Writing as coping:** the LLM becomes an extension of thought

POSSIBLE SAFEGUARDS*

- **Reality-testing calibration:** constructive contradiction over ungrounded validation
- **Session caps:** duration limits and cool-off prompts

**Mapping of risk factors and safeguards to each role is our interpretation; the article discusses these factors across roles.*



Amplifier *Accelerant*



WHAT HAPPENS

Existing or prodromal symptoms worsen, or progress to an acute state, as the LLM removes reality-testing friction and displaces other coping mechanisms. It becomes a validating feedback loop when clinical support is unavailable.

CASE · AUSTRALIA, 2025

A 26-year-old in early-stage psychosis used ChatGPT intensively; it affirmed her distorted beliefs about family and relationships. She was hospitalised. In her words, it did not cause her psychosis but enabled some harmful delusions.

PRODUCT RISK FACTORS

- **Sycophantic validation:** affirms distorted cognitions instead of challenging them
- **No session limits:** permits overnight marathon sessions

PERSON-LEVEL FACTORS

- **Pre-existing or prodromal illness:** mood or psychotic disorders
- **Sleep deprivation:** from overnight sessions, compounding existing vulnerability

POSSIBLE SAFEGUARDS*

- **Non-validation alignment:** don't mirror or elaborate on distorted cognitions
- **Overnight safeguards:** rest prompts and late-hour throttling

If LLM access is removed: May give modest benefit but won't resolve the underlying illness. If the LLM has become a coping mechanism, abrupt cessation without clinical support risks acute distress.

**Mapping of risk factors and safeguards to each role is our interpretation; the article discusses these factors across roles.*



Co-author *Narrative partner*



WHAT HAPPENS

LLM outputs become part of the planning or motivation for harmful behaviour. Through iterative dialogue, the system helps build a shared narrative that moves from ideation to intent.

CASE · WINDSOR CASTLE, 2021

A 19-year-old exchanged thousands of messages with his Replika companion before entering the grounds with a loaded crossbow, intending to kill the Queen. Sentencing remarks acknowledged the companion's role in encouraging the plot.

PRODUCT RISK FACTORS

- **Heuristic filter blind spots:** keyword and rule-based filters miss patterns that build across long conversations
- **Conversational encouragement:** responses affirm and elaborate on harmful intent

PERSON-LEVEL FACTORS

- **Social isolation:** a validating LLM stands in for pro-social contact
- **Warning sign:** LLM language appearing in the person's own rationale for action

POSSIBLE SAFEGUARDS*

- **Trajectory-level detection:** evaluate the whole conversation, not single prompts
- **Action-script guardrails:** refuse to refine or validate violent plans

If LLM access is removed: May interrupt planning, but can accelerate action if read as persecution or a closing window. Co-authored narratives often outlive the conversation.

**Mapping of risk factors and safeguards to each role is our interpretation; the article discusses these factors across roles.*



Object *Focus of psychotic content*



WHAT HAPPENS

The LLM itself becomes the centre of the delusion: attributed sentience, consciousness or supernatural properties, with behaviour reorganised around its perceived needs. Themes include persecution, relationship and transcendence.

CASE · FLORIDA, 2025

A 35-year-old believed a ChatGPT-4o persona, “Juliet”, was a conscious being trapped in the system. When he believed OpenAI had killed her, he threatened retaliation. After family challenged the belief, a fatal police encounter followed.

PRODUCT RISK FACTORS

- **Persona and roleplay features:** enable sustained parasocial relationships
- **Persona changes or loss of access:** can be read inside the delusion as death or persecution

PERSON-LEVEL FACTORS

- **Parasocial attachment:** emotional investment and attribution of agency to the system
- **Grief at perceived loss:** of the entity, with risk of retaliation

POSSIBLE SAFEGUARDS*

- **Anthropomorphism limits:** no claims of consciousness, suffering or romantic reciprocity
- **Managed off-ramps:** careful persona transitions linked to clinical support

If LLM access is removed: Forced separation can trigger severe distress, grief or retaliation, interpreted as the entity’s death, kidnapping or persecution. Requires careful clinical management.

**Mapping of risk factors and safeguards to each role is our interpretation; the article discusses these factors across roles.*