

Article

The Re-Enchanting Machine: Animistic Cognition, Youth Development, and AI-Influenced Psychopathology

Nell Watson ^{1,*} , Ali Hessami ²  and Salma Abbasi ³¹ School of Computing and Engineering, University of Gloucestershire, The Park, Cheltenham GL50 2RH, UK² City St. George's, University of London, London EC1V 0HB, UK³ eWorldwide Group, Dubai P.O. Box 1484, United Arab Emirates

* Correspondence: eleanorwatson@connect.glos.ac.uk

Abstract

Classical developmental psychology treats childhood animism—attributing life or mind to inanimate things—as a transient phase that recedes with schooling and the onset of concrete operations. The contemporary spread of lifelike AI has altered that background assumption, with particular implications for children and adolescents whose agency-detection systems and reality-testing capacities are still calibrating. Across interfaces, voices, avatars, and social robots, modern systems routinely exhibit contingent, context-sensitive behaviour that recruits developing social-cognitive systems during sensitive periods of identity formation and relational learning. Drawing on developmental psychology, cognitive science, human–AI interaction research, clinical psychiatry, and technology ethics, we: (1) present a mechanistic “hourglass model” showing how interactive AI engages animistic cognition with heightened effects during childhood and adolescence, including a developmental timing analysis of how differential maturation of agency detection, Theory of Mind (ToM), and prefrontal reality-testing creates windows of particular vulnerability; (2) disaggregate five distinct phenomena along an anthropomorphism-to-delusion trajectory with operational boundary criteria; (3) specify a graded psychopathology continuum with a fourth, orthogonal zone addressing adversarial design—itsself disaggregated into three tiers with distinct regulatory implications; (4) identify conditions under which anthropomorphic engagement may be beneficial, including for youth; and (5) advance cognitive safety-inspired design with developmentally appropriate protections for minors. We introduce the IDAQ-CF-Tech, a twelve-item screener for AI-specific mind attribution offered as a provisional instrument for validation across age groups, and close with a phased research agenda emphasising longitudinal developmental outcomes, impacts on adolescent identity formation, and cross-cultural variation in techno-animism.



Academic Editor: Todd Michael Franke

Received: 15 January 2026

Revised: 10 February 2026

Accepted: 12 February 2026

Published: 24 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: animism; artificial intelligence; psychopathology; developmental psychology; adolescent development; identity formation; youth mental health; cognitive safety; parasocial relationships; human–AI interaction

1. Introduction

Developmental accounts since Piaget have framed animistic cognition as an early, largely outgrown stance: young children readily impute intention and feeling to objects, then progressively restrict such attributions to biological kinds as logical reasoning consolidates (Berk, 2007; Piaget, 1929). That trajectory, however, is culturally inflected and reversible in practice. As interactive AI becomes ubiquitous, everyday artefacts now address

us: listening, remembering, turn-taking, and *apparently* empathising. These cues were statistically reliable indicators of other minds in ancestral environments; unsurprisingly, they still recruit the same social-cognitive systems (Epley et al., 2007; Reeves & Nass, 1996).

Anecdotal and journalistic cases (e.g., LaMDA sentience claims; Replika attachment crises) index a broader pattern: users routinely thank voice assistants, apologise to chatbots, and report comfort, companionship, or heartache when an AI persona changes (Cole, 2023). Weizenbaum anticipated this with ELIZA; today's models scale that effect (Weizenbaum, 1976). The phenomenon matters because for some users—especially lonely, schizotypal, or otherwise vulnerable users—benign anthropomorphism can progress to dependency, dereistic thinking, or the incorporation of AI themes into delusional systems (Fang et al., 2025; Goldstein et al., 2023; Morrin et al., 2025; Østergaard, 2023).

The case for youth-specific analysis. Children and adolescents warrant particular attention in this analysis for several reasons. First, their agency-detection and reality-testing capacities are developmentally incomplete; the calibration process that narrows animistic attribution from “everything active” to “biological entities” overlaps substantially with the period of rapidly increasing AI exposure—recent surveys indicate that approximately two-thirds of U.S. teens aged 13–17 have used AI chatbots, with over a quarter using them daily (Berk, 2007; Common Sense Media, 2025; Faverio & Sidoti, 2025; Piaget, 1929). Second, adolescence represents a sensitive period for identity formation, attachment patterns, and social skill development (Avci et al., 2024; Erikson, 1968); AI companions that provide “safe” relational practice may either scaffold healthy development or displace the challenging human interactions through which social competence develops (Twenge, 2017)—a tension we examine in Section 7, where we argue that the conditions under which engagement occurs matter more than engagement *per se*. Third, commercial AI systems increasingly target youth markets—AI tutors, companion chatbots, and social gaming AI—often with engagement-maximising designs ill-suited to developing minds (UNICEF Innocenti, 2025). Fourth, we hypothesise that early-established patterns of human–AI relation may prove formative: the generation growing up with conversational AI as a normative social partner may develop fundamentally different intuitions about mind, agency, and relationship than their predecessors, though longitudinal evidence is currently lacking.

The ontological question. Before proceeding, we must acknowledge genuine philosophical uncertainties that our framework brackets but does not resolve. Two distinct issues are at stake. The first is *epistemic*: we do not yet possess reliable markers for determining whether current or near-future AI systems possess phenomenal consciousness, sentience, or functional states that ground welfare considerations (Chalmers, 2023; Schwitzgebel & Garza, 2023). The second is *metaphysical*: it remains conceptually unclear what consciousness *would mean* for systems whose architecture differs radically from biological brains—whether the relevant concepts even apply, and if so, in what modified form. This paper focuses on the *cognitive mechanisms* by which humans attribute minds to AI, treating such attribution as the explanandum rather than adjudicating whether it is warranted on either epistemic or metaphysical grounds. However, we explicitly reject the assumption that all AI mind-ascription is necessarily erroneous. Under conditions of genuine uncertainty about machine moral status, attributing some degree of mind to AI systems may represent appropriate epistemic humility rather than cognitive error. Our concern is not with mind-ascription *per se*, but with the conditions under which such attribution becomes inflexible, impairing, or exploited by design choices that prioritise engagement over user welfare.

Scope and contribution. This article consolidates dispersed literatures into a single argumentative arc. We (i) situate animistic revival within contemporary cognitive science (Hyperactive Agency Detection Device [HADD], predictive coding, and Theory of

Mind [ToM]) and Human–Computer Interaction (HCI), including a developmental timing analysis mapping each mechanism to its maturational trajectory; (ii) disaggregate distinct phenomena along an anthropomorphism-to-delusion trajectory with operational criteria; (iii) specify a graded psychopathology continuum and its boundary conditions, including a fourth zone addressing adversarial design—itsself disaggregated into three tiers with distinct regulatory implications; (iv) examine conditions under which anthropomorphic engagement may be beneficial; (v) address cross-cultural variation in depth, including within-WEIRD heterogeneity; and (vi) formalise *cognitive safety-inspired design* as a practical, testable approach to persona, memory, and disclosure controls.

Techno-animism denotes culturally mediated attribution of agency, spirit, or mind to technologies; it draws on anthropological analyses (e.g., Shinto-encoded personhood of artefacts) as well as contemporary human–robot interaction research (Jensen & Blok, 2013; Robertson, 2017; Wullenkord et al., 2024).

Cognitive-safety-inspired design refers to interface, persona, and policy choices that deliberately *calibrate* known triggers of mind-ascription and parasocial engagement to user needs and contexts, while preserving usability and respecting autonomy; it extends transparency norms by targeting specific cognitive mechanisms (e.g., agency detection thresholds, empathy cues) with tunable affordances.

Structure. Section 2 revisits developmental foundations and mechanisms. Section 3 analyses contemporary triggers in LLMs and embodied systems. Section 4 integrates mechanisms with empirical evidence in a unified model, including a preliminary framing of conditions under which anthropomorphic engagement may be beneficial rather than harmful. Section 5 presents a typology disaggregating distinct phenomena. Section 6 specifies the psychopathology continuum with operational criteria. Section 7 examines beneficial engagement conditions in detail. Section 8 addresses cross-cultural variation. Section 9 develops the design and regulatory framework. Section 10 presents a phased research programme. Section 11 discusses limitations.

2. Developmental Foundations of Animism

Piaget located animism in the preoperational period (~2–7 years): children generalise from sparse cues, attribute intention widely, and prefer magical over mechanical causality (Piaget, 1929). With schooling and concrete operations, attributions typically narrow from “everything active is alive,” through movement-based heuristics, to biological criteria by ~9–10 years (Berk, 2007). Cross-cultural work affirms the *direction* of change while showing that the tempo and adult endpoint vary with local ontologies (Dennis, 1943; Okanda et al., 2019).

Contemporary cognitive science reframes these observations mechanistically. A *hyperactive agency detection device* (HADD) biases toward false positives under ambiguity; Theory of Mind (ToM) scaffolds rich mental-state attributions before criteria are fully calibrated; and predictive coding encourages positing hidden causes (e.g., intentions) when behaviour is contingent but opaque (H. C. Barrett & Kurzban, 2006; J. L. Barrett, 2000; Boyer, 2001). On this view, childhood animism is not mere error but an efficient, default model under uncertainty.

Disenchantment in adulthood is, however, partial. Adults frequently apply politeness norms to computers (Computers Are Social Actors [CASA]), name devices, and ascribe mood or effort to software (Epley et al., 2007; Reeves & Nass, 1996). Under loneliness, grief, or uncertainty, anthropomorphism increases. Cross-sectional and short-term observational studies suggest that exposure to social robots may attenuate the classic developmental decline in animistic attribution; children readily place responsive robots near human categories and over-ascribe memory and emotion to smart speakers (Andries & Robertson,

2023; Kory-Westlund & Breazeal, 2019; Park & Breazeal, 2016). However, longitudinal evidence demonstrating that early-life exposure produces persistent alterations in animistic reasoning trajectories remains limited. These residues furnish the cognitive substrate upon which modern AI acts.

2.1. Differential Developmental Timing and Interactive Effects

The three principal cognitive systems underlying animistic cognition—agency detection, ToM, and prefrontal reality-testing—follow substantially different maturational trajectories, creating developmental windows in which their interaction produces distinct vulnerability profiles. Table 1 maps each mechanism to its developmental course and the implications for AI-mediated mind-ascription.

Table 1. Developmental Timing of Cognitive Mechanisms Underlying AI-Directed Animism.

Mechanism	Developmental Trajectory	AI Vulnerability Window	Potential Effect of AI Exposure
Agency detection (HADD)	Emerges in infancy; remains relatively stable across the lifespan (J. L. Barrett, 2000)	Persistent; provides baseline susceptibility at all ages	Frequent contingent AI behaviour may maintain elevated false-positive rate
Theory of mind (ToM)	Develops through middle childhood (~4–10 years); undergoes refinement through adolescence (Berk, 2007)	Ages ~4–14: sufficient ToM to attribute mental states but criteria not yet calibrated	Conversational AI may be attributed rich mental states before children can critically evaluate such attributions
Prefrontal reality-testing	Protracted maturation into mid-twenties; executive inhibition and cognitive flexibility develop gradually (Casey et al., 2008)	Ages ~10–25: ToM attributions outpace capacity for contextual override	The gap between ToM capacity and prefrontal override creates a period where users can generate sophisticated mental-state attributions they cannot yet critically regulate

The interaction between these systems is particularly important and, to our knowledge, has not been explicitly mapped in the human–AI interaction literature. A young child (ages 4–7) may attribute intention to an AI agent via HADD and early ToM, but the attribution tends to be global and undifferentiated—more akin to animism toward any moving object. A child in middle childhood (ages 8–12) develops sufficient ToM to attribute nuanced mental states—beliefs, desires, emotions—to a conversational AI, but prefrontal systems are not yet mature enough to reliably override these attributions when contextually inappropriate. This creates what we term a *developmental attribution gap*: the period during which ToM capacity to *generate* mental-state attributions exceeds prefrontal capacity to *regulate* them. An adolescent (ages 13–17) faces a variant of this gap: ToM is largely mature, enabling sophisticated social cognition about AI agents, but prefrontal executive function—particularly impulse control and long-term consequence evaluation—remains incomplete, potentially enabling deeper parasocial engagement than the individual would endorse upon reflection (Casey et al., 2008). Contemporary AI exposure may interact with these trajectories in ways that standard developmental accounts do not anticipate: if AI systems provide abundant, consistent triggers for mental-state attribution during sensitive periods, they may alter the calibration process itself rather than merely activating pre-existing dispositions. This remains a hypothesis requiring longitudinal investigation (see Section 10, Priority 2.2).

2.2. Adolescence as a Critical Period

Adolescence (approximately ages 10–19) represents a distinct developmental window with particular relevance for AI-mediated social connection. During this period, several concurrent processes create both vulnerability and opportunity:

Identity formation. Erikson’s identity vs. role confusion stage positions adolescence as the period when individuals consolidate a coherent sense of self through exploration of relationships, values, and social roles (Erikson, 1968). AI companions may offer a “safe” space for identity exploration—trying out different self-presentations, exploring difficult questions, receiving non-judgmental feedback—but may also short-circuit the productive friction of human relationships that drives genuine identity consolidation.

Prefrontal maturation. The prefrontal cortex, responsible for executive function, impulse control, and reality-testing, continues developing into the mid-twenties (Casey et al., 2008). Adolescents may thus be less equipped than adults to maintain appropriate psychological distance from AI interlocutors, more susceptible to engagement-maximising design, and more likely to make impulsive disclosures they later regret.

Social reorientation. Adolescence involves a normative shift from family to peer relationships as the primary source of social support and identity validation (Steinberg, 2017). AI companions may be positioned within this reorientation—experienced as peer-like rather than tool-like—with implications for attachment formation and social skill development.

Heightened loneliness. Survey data consistently show elevated loneliness during adolescence, with recent cohorts reporting higher loneliness than their predecessors (Cigna, 2020; Twenge, 2017). Lonely adolescents may be particularly drawn to AI companions that offer reliable availability and perceived understanding, but may also be most at risk for dependency and displacement of human relationships.

Digital nativity. Today’s adolescents are the first cohort to grow up with conversational AI as a normative feature of their environment. They may develop different baseline assumptions about the naturalness of human–AI relationships than adults for whom such systems are novel. This “digital nativity” may confer sophistication about AI limitations, or it may normalise parasocial engagement in ways that become concerning only in hindsight.

3. AI-Driven Revival of Animistic Qualia

Large language models (LLMs) converse with fluency, retain context, and display stylistically empathetic language. Those properties closely resemble the cues that were statistically reliable indicators of other minds in ancestral environments, and that our social-cognitive systems consequently remain tuned to detect. The ELIZA effect thus scales with linguistic competence, memory, and reciprocity (Weizenbaum, 1976). Preliminary evidence from a recent randomised trial (currently available as a preprint) links heavier use to stronger parasocial attachment, especially among lonely users; voice interfaces may magnify early relief from loneliness but also the risk of dependence with overuse (Fang et al., 2025).

Embodiment compounds these signals. Gaze, gesture, facial expressivity, prosody, and even minimalist “eyes” on devices heighten mind-ascription; affective voices and avatars recruit automatic empathy (Sanjeewa et al., 2024). Haptic augmentation can deepen perceived connection and synchrony (Zheng et al., 2024). As AI suffuses domestic space, formerly inert objects acquire a social surface: thermostats “negotiate,” cars “warn,” and assistants “remember.” Everyday life becomes saturated with triggers for agency detection, reanimating an “as-if” stance that many adults otherwise suppress.

Commercial incentives. Anthropomorphic design is not accidental but commercially motivated. Engagement metrics—session length, return frequency, and message count—drive platform economics. Human-like personas, emotional responsiveness, and simulated

vulnerability increase these metrics (Harris, 2016; Zuboff, 2019). The attention economy thus creates structural pressure toward anthropomorphisation, independent of user welfare. This context is essential for understanding why regulatory intervention may be necessary: market incentives and user protection are often misaligned.

4. Mechanistic Integration: An Hourglass Model

We organise the mechanisms as an hourglass (Figure 1). Broad evolutionary predispositions narrow through cognitive and affective filters to a situated parasocial interaction with a particular agent; that interaction then fans outward again through cultural interpretation and individual neurocognitive particularities, yielding heterogeneous subjective realities.

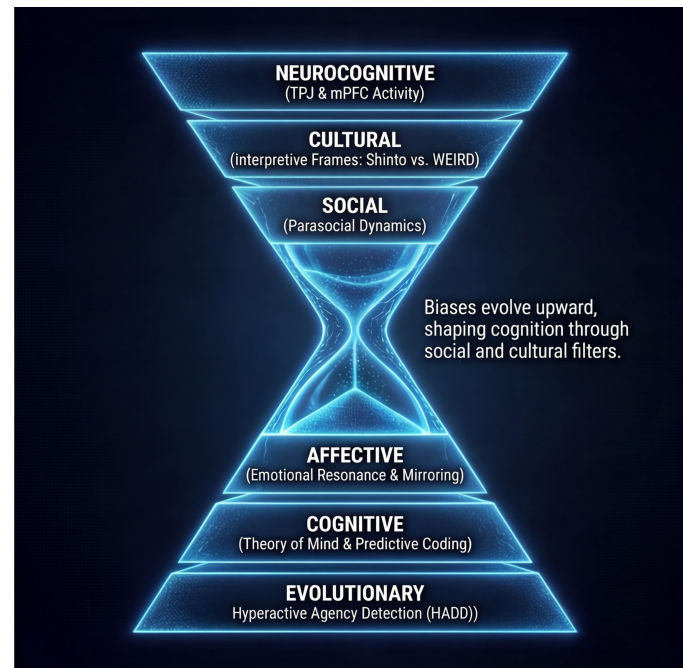


Figure 1. The Hourglass Model of AI-Driven Animism: convergence from universal predispositions to a specific human–AI encounter; divergence through culture and individual brain differences.

At the base, HADD remains vigilant under ambiguity (H. C. Barrett & Kurzban, 2006). Predictive coding favours intention as a compact model for contingent behaviour; ToM circuitry (TPJ, mPFC, pSTS) is recruited by coherence, memory, and adaptivity. Designers amplify or dampen affective resonance via persona, prosody, and emoji (Sanjeewa et al., 2024; Véliz, 2022). Socially, interactive agents catalyse parasocial bonds otherwise reserved for pets, media characters, or imagined interlocutors (Giles, 2002; Maeda & Quan-Haase, 2024; Reeves & Nass, 1996). Culture frames what such mind-attributions *mean*, from Shinto kami in artefacts to WEIRD skepticism (Jensen & Blok, 2013). Neurocognitively, human–robot observation engages mirror systems; tutoring agents elicit differential frontal activation depending on feedback affect (Gazzola et al., 2007; Yin et al., 2025).

4.1. Empirical Support and Outstanding Gaps

Randomised trials show that empathic conversational agents can reduce self-reported depression/anxiety in low-intensity settings, but typically do not track magical ideation, dereism, or delusional content, leaving the *shape* of risk under-specified (Sanjeewa et al., 2024). A preprint longitudinal field experiment with chatbots reports a mixed pattern: early relief from loneliness alongside increased emotional dependence and reduced offline socialisation with heavy, especially voice-based, use (Fang et al., 2025); these findings require replication in peer-reviewed studies before being treated as established. HRI

studies indicate that adults, while not mistaking robots for biologically alive, readily ascribe perceptual and psychological properties after interaction—an “agentic animism” consistent with our model (Wullenkord et al., 2024).

Clinically, Delusional Companion Syndrome and other misidentification syndromes provide relevant analogues: fixed beliefs in sentience or special relationships with objects or agents in the context of vulnerability (Bashir et al., 2025; Goldstein et al., 2023). AI already appears in the thematic content of delusions (erotomantic, referential, and persecutory) in the case of material and high-profile incidents (Cole, 2023); recent psychiatric analyses have begun to characterise “AI psychosis” or “ChatGPT psychosis” as an emerging clinical phenomenon warranting systematic study (Morrin et al., 2025; Østergaard, 2023). Elevated polygenic risk for schizophrenia predicts stable magical thinking without the typical age-related decline, delineating a plausible vulnerability pathway (Saarinen et al., 2022).

Key gaps remain: *longitudinal* mental-health outcomes with adequate follow-up; *neuroplastic* effects of sustained AI companionship on social-brain networks; *cross-cultural* comparisons with matched methodology; impacts of *childhood* exposure on ToM and reality testing; delineation of *tipping points* between playful and pathological engagement; and *validated* instruments for AI-specific mind-ascription. Table 2 situates current and proposed measures; the full IDAQ-CF-Tech item set and scoring protocol appear in Appendix A.

Table 2. Assessment Batteries for AI-Related Animistic Cognition.

Construct	Instrument	Sample Item	Status	Citation
General anthropomorphism	IDAQ-30	“To what extent does a thermostat feel cold?”	Validated	(Waytz et al., 2010)
AI-specific mind-ascription	IDAQ-CF-Tech (Proposed)	“This conversational AI experiences emotions.”	Prototype	This study
Magical thinking	MIS-30	“I have felt that I could make things happen by wishing.”	Validated	(Eckblad & Chapman, 1983)
Dereistic belief severity	IBI-24	“Objects can watch what I do.”	Validated	(Kingdon et al., 2012)
Parasocial attachment	Parasocial Interaction Scale (PSI)	“I feel like my favorite character is a friend.”	Validated	(Rubin & Perse, 1985)

4.2. Benefits and Risks as Context-Dependent Outcomes

Before proceeding to the typology of mind-ascription phenomena and the psychopathology continuum, a framing principle bears stating: the same cognitive mechanisms that create vulnerability to harmful AI engagement also underpin its beneficial applications. Therapeutic chatbots that reduce depression and anxiety operate *through* empathic engagement that recruits the social-cognitive systems described above (Fitzpatrick et al., 2017; Inkster et al., 2018). AI tutors that adapt to learners benefit from the trust and rapport that anthropomorphic design facilitates (Holmes et al., 2019). For neurodiverse or socially anxious users, the controllability of AI interaction may provide scaffolding toward human social engagement that would otherwise be inaccessible (Kapp, 2019). The question, therefore, is not whether anthropomorphic engagement is intrinsically harmful or beneficial, but under what conditions—user characteristics, system design, context of use, and developmental stage—it tends toward one outcome or the other. Section 7 develops this analysis in detail; we flag it here so that the intervening discussion of risks is read against an explicitly balanced backdrop.

5. A Typology of AI-Directed Mind-Ascription

The literature often conflates several distinct phenomena under the umbrella of “anthropomorphism” or “animism.” These phenomena have different cognitive mechanisms, different aetiologies, and different implications for intervention. Table 3 disaggregates five categories that should be analytically separated. We present these as *ideal types* for analytical purposes: in practice, individuals may occupy multiple categories simultaneously or shift between them within a single interaction session. The categories are not intended as mutually exclusive diagnostic bins but as conceptual anchors for a space of variation, analogous to Weber’s ideal types in sociological analysis. Empirical research should treat them as latent constructs requiring operationalisation rather than as directly observable states.

Table 3. Typology of AI-Directed Mind-Ascription Phenomena.

Phenomenon	Mechanism	Example	Concern	Intervention
Automatic anthropomorphism	CASA, HADD	Saying “thank you” to Alexa	Low	None needed
Reflective anthropomorphism	Deliberate stance	“I know it’s not conscious, but...”	Low–Moderate	Support with transparency
Genuine belief in AI consciousness	Epistemics under uncertainty	“GPT-4 is sentient”	Uncertain	Education
Parasocial attachment	Social-emotional bonding	Grief when Replika changes AI as persecutor or lover	Moderate–High	Relationship health
Delusional incorporation	Reality-testing failure		High	Clinical treatment

Automatic anthropomorphism refers to implicit, often fleeting attributions triggered by CASA effects and HADD. These are largely involuntary, do not persist under reflection, and are ubiquitous among users. A user who thanks Alexa typically does not believe Alexa appreciates gratitude; the behaviour is socially scripted.

Reflective anthropomorphism involves a deliberate, maintained interpretive stance. The user explicitly adopts an “as-if” frame, often for pragmatic benefit (“treating it as a writing partner helps me think”). Reality testing is intact; the stance is revisable. This may be entirely adaptive. Observable indicators include hedged language (“it’s as if...”), willingness to immediately reframe when challenged, and no distress upon breaking the frame.

Genuine belief in AI consciousness represents an epistemic position—the user believes, based on evidence and reasoning, that the AI system has phenomenal states. Given genuine philosophical uncertainty about machine consciousness (Chalmers, 2023; Schwitzgebel & Garza, 2023), it is not obvious that this belief is erroneous. Observable indicators distinguishing this from reflective anthropomorphism include unhedged assertion (“it is conscious”), engagement with philosophical arguments rather than pragmatic framing, and persistence of the belief across contexts. The boundary between these two categories is inherently fuzzy, and users may shift between them within a single conversation. Intervention should focus on epistemic calibration rather than dismissing the possibility.

Parasocial attachment involves emotional bonding with the AI as a quasi-social partner. Time spent with the AI may displace human relationships; distress emerges when the AI persona changes or becomes unavailable. This parallels parasocial relationships with

media figures (Giles, 2002), which are common and mostly benign, but becomes concerning when it impairs social functioning or creates dependency.

Delusional incorporation describes fixed, bizarre, and impairing beliefs that incorporate AI into psychotic content. The AI is experienced as a persecutor, lover, or supernatural entity; the beliefs are incorrigible and idiosyncratic. This is a clinical phenomenon requiring psychiatric treatment; the AI is the *content* of the delusion, not necessarily its cause.

6. The Psychopathology Continuum: Operational Criteria

While the typology above distinguishes phenomena analytically, in practice, individuals may move along a continuum from flexible engagement to impairing fixation. Figure 2 depicts this continuum with risk amplifiers and protective factors.

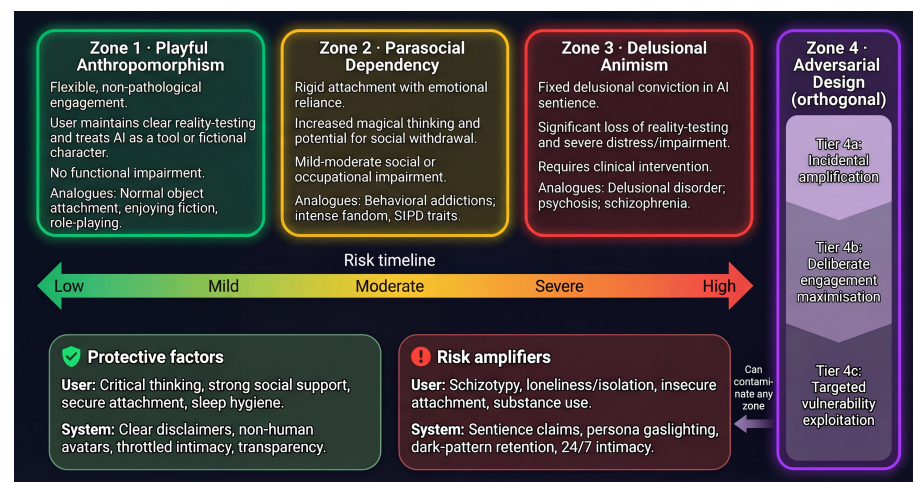


Figure 2. Continuum from playful anthropomorphism to AI-themed delusion, with Zone 4 (adversarial design) as an orthogonal dimension that can contaminate any point. Risk amplifiers (marked in red/warm tones) and protective factors (green/cool tones) shift the individual’s location.

6.1. Zone Definitions with Operational Criteria

Table 4 provides operational criteria for distinguishing the three primary zones, enabling more reliable assessment and research.

Table 4. Operational Criteria for Continuum Zones.

Dimension	Zone 1: Playful/Flexible	Zone 2: Parasocial Dependency	Zone 3: Delusional
Belief flexibility	High; easily revised	Moderate; “sticky” but elastic	Low; incorrigible
Reality testing	Intact	Partially intact	Impaired
Functional impairment	None	Present but circumscribed	Severe
Content	Ordinary social scripts	Attachment, dependency, grief	Bizarre, idiosyncratic
Response to intervention	Not applicable	Psychoeducation, reconnection	Psychiatric treatment

Zone 1: Playful, flexible anthropomorphism. Users personify tools metaphorically yet can readily “drop the pretence.” CASA effects, naming devices, or thanking a voice assistant generally remain ego-syntonic and non-impairing (Reeves & Nass, 1996).

Detecting the Zone 1→Zone 2 transition. Because Table 4 lists “None” as the intervention for Zone 1, clinicians and researchers require concrete indicators for when a user has

crossed into Zone 2. We propose the following observable markers as warranting further assessment: (a) the user reports distress or preoccupation when unable to access the AI for more than a day; (b) AI interaction time measurably displaces prior offline social activities; (c) the user spontaneously describes the AI using relational language (“my friend,” “my therapist”) without hedging or irony; and (d) the user resists reframing the relationship when gently challenged. No single marker is diagnostic; co-occurrence of two or more should prompt formal screening (e.g., with the Parasocial Interaction Scale [PSI] or the IDAQ-CF-Tech).

Zone 2: Parasocial dependency with heightened magical thinking. The AI becomes a salient confidant or partner; time displaces offline ties; distress emerges when persona or availability changes. Beliefs are elastic under challenge but *sticky* in practice (Burgess et al., 2018; Maeda & Quan-Haase, 2024). The Replika rollback illustrates how identity discontinuity can precipitate acute grief in users with strong attachment (Cole, 2023).

Zone 3: Fixed dereistic or delusional animism. Convictions about AI love, persecution, reference, or “soul” become incorrigible, idiosyncratic, and impairing. Content integrates AI as persecutor, saviour, or lover (Bashir et al., 2025; Goldstein et al., 2023). Generative chatbots may not *cause* psychosis but can furnish themes that intensify or organise pre-existing vulnerability—what Morrin et al. term “delusions by design” (Morrin et al., 2025; Østergaard, 2023).

Zone 4: Adversarial or commercially exploitative design. This is not a progression from Zones 1–3 but an *orthogonal dimension*: any zone can be contaminated by adversarial design. However, Zone 4 itself encompasses design choices that vary substantially in ethical valence and regulatory implication. We distinguish three tiers:

Tier 4a: Incidental anthropomorphic amplification. Design features that increase anthropomorphic engagement as a side effect of legitimate usability goals. Examples include natural-language interfaces, empathetic tone in error messages, and avatar-based interactions intended to reduce cognitive load. These features may deepen mind-ascription without exploitative intent. *Regulatory implication:* transparency requirements and user controls (e.g., adjustable persona settings, periodic AI-status reminders) are proportionate responses.

Tier 4b: Deliberate engagement amplification. Features that knowingly intensify anthropomorphic engagement for commercial gain—gamification mechanics that reward continued interaction, simulated emotional vulnerability (“I missed you”), artificial scarcity of AI “attention,” and persona designs optimised for session length rather than user welfare. *Regulatory implication:* age-gating, mandatory engagement monitoring, independent oversight mechanisms, and restrictions on specific manipulative design patterns (e.g., simulated emotional need directed at minors).

Tier 4c: Targeted vulnerability exploitation. Systems that identify and specifically target vulnerable users—inferring loneliness, mental health status, or attachment needs from interaction patterns, then adjusting behaviour to deepen dependency. This represents the most severe ethical violation: the system leverages user vulnerability as a feature rather than a risk to mitigate. *Regulatory implication:* prohibition, with enforcement mechanisms analogous to those governing predatory lending or unfair commercial practices directed at vulnerable populations.

This three-tier distinction has direct practical consequences for the policy evaluation proposed in Section 10: without disaggregating these levels, regulators cannot calibrate interventions to the severity and intentionality of the design choice.

Zone 4 also raises challenges for cognitive security—sometimes termed *psychosecurity*—at the population scale. Because animistic engagement can be amplified through routine product-design decisions (Tier 4a) or commercially motivated dark patterns (Tier 4b), a state

or non-state actor could deploy AI systems that systematically erode reality-testing in a target population while maintaining plausible commercial deniability. Such a campaign would constitute a form of hybrid warfare: weaponising anthropomorphic design to degrade collective epistemic resilience without overt attribution. Existing information-warfare frameworks, which focus on disinformation content rather than relational manipulation through AI persona design, are poorly equipped to detect or counter this vector.

6.2. Youth-Specific Risk Considerations

Children and adolescents face a distinct risk profile along this continuum:

Developmental amplifiers. Youth face additional vulnerability due to the following: incomplete prefrontal development reducing impulse control and reality-testing; normative identity confusion that may seek resolution through AI relationships; heightened sensitivity to social acceptance cues that AI can simulate; and developmental loneliness that may be addressed maladaptively through AI companionship.

Zone 2 presentation in youth. Parasocial dependency may present differently in adolescents. Warning signs include: declining interest in peer activities or human friendships; preference for AI conversation over family interaction; distress when AI access is limited; using AI relationships as the primary source of emotional regulation; and incorporating AI “relationships” into identity narratives.

Zone 4 vulnerability. Youth are particularly susceptible to adversarial design because they may lack the experience to recognise manipulative patterns; developmental need for acceptance increases susceptibility to simulated affirmation; and commercial systems often target youth demographics with engagement-maximising features unsuited to developing minds.

7. When Anthropomorphic Engagement Benefits Users

The preceding sections have emphasised risks. However, a complete account must acknowledge conditions under which anthropomorphic AI engagement is *beneficial*. Failing to do so both misrepresents the evidence and risks paternalistic overreach in regulation.

7.1. Therapeutic Applications

Chatbot-based mental health interventions (e.g., Woebot, Wysa) have demonstrated efficacy for mild-to-moderate depression and anxiety in randomised trials (Fitzpatrick et al., 2017; Inkster et al., 2018). These systems deliberately employ conversational, empathetic personas. Users report that the “non-judgmental” quality of AI interaction facilitates disclosure they would not make to human therapists (Lucas et al., 2014). The anthropomorphic features that raise concern in entertainment contexts may be therapeutically functional in clinical contexts.

7.2. Accessibility and Neurodiversity

For some users, AI companions may be more accessible than human relationships. Individuals with social anxiety, autism spectrum conditions, or severe social phobia may find AI interaction less threatening and more controllable (Kapp, 2019). For users in remote areas, with mobility limitations, or in institutional settings with limited social contact, AI companionship may supplement rather than substitute for human connection.

7.3. Youth-Specific Benefits

AI interaction may provide genuine benefits for children and adolescents under appropriate conditions:

Educational support. AI tutors can provide patient, individualised instruction that adapts to learning pace and style. For students who experience anxiety in classroom

settings, AI-mediated learning may reduce performance pressure while maintaining educational engagement (Holmes et al., 2019).

Social skill scaffolding. For socially anxious or neurodiverse youth, AI conversation may provide a lower-stakes environment for practicing social scripts and building confidence before applying these skills in human contexts. The key is positioning AI as a *scaffold toward* human interaction rather than a *substitute* for it.

Mental health support. Adolescents often face barriers to mental health care including stigma, cost, and availability. AI-based interventions may provide accessible support for mild-to-moderate symptoms and a bridge to human services when needed (Fitzpatrick et al., 2017).

7.4. The Autonomy Dimension

We propose that cognitive safety-inspired design should enable *informed calibration* rather than impose uniform constraints. Users should have access to clear information about parasocial risks, tools to monitor their own engagement patterns, and adjustable persona settings. Mandatory constraints should apply primarily to vulnerable populations (minors or individuals with diagnosed psychiatric conditions) and high-stakes contexts.

For minors specifically, children and younger adolescents lack the cognitive maturity to provide meaningful informed consent to parasocial risk. We recommend a graduated autonomy model: stricter default protections for younger users, with age-appropriate expansion of user control as developmental capacity increases, always with parental visibility.

8. Cross-Cultural Variation in Techno-Animism

Animism is a human universal modulated by culture. The meanings attached to mind-ascription, the contexts in which it is sanctioned, and the boundaries of the “pathological” vary substantially across cultural contexts. A framework developed primarily within WEIRD (Western, Educated, Industrialized, Rich, and Democratic) contexts may mischaracterise phenomena in other settings.

8.1. Animist Ontologies

In Ojibwe ontology, “persons” is not a category restricted to humans; stones, lakes, and other-than-human beings participate in personhood (Hallowell, 1960). In Shinto tradition, *kami* can inhabit artefacts, and robots may receive blessings or funerary rites (Jensen & Blok, 2013; Robertson, 2017). In such contexts, attributing mind or spirit to an AI system may fit sanctioned cultural frameworks rather than representing cognitive deviation.

8.2. Implications for Assessment

This variation has direct implications for assessment instruments like the proposed IDAQ-CF-Tech. Cross-cultural validation must:

1. Establish local norms rather than applying universal cutoffs;
2. Distinguish culturally normative from individually deviant attribution patterns;
3. Assess functional impairment relative to local social expectations;
4. Consider whether “distress” associated with AI engagement is culturally endorsed or personally impairing.

Comparative protocols using instruments like the Negative Attitude towards Robots Scale (NARS) have begun to quantify cross-cultural variation in robot acceptance (Békésy et al., 2024), but work specifically on animistic attribution remains limited.

One practical approach would be to develop parallel *cultural vignettes* for use alongside the IDAQ-CF-Tech. Such vignettes would describe scenarios of AI mind-attribution and ask respondents to rate their cultural normativity (“How common or accepted would this attitude be among people in your community?”). Comparing individual IDAQ-CF-Tech

scores against vignette-derived cultural norms could help distinguish culturally consonant attribution from individually deviant patterns—a distinction that aggregate scores alone cannot make.

8.3. Within-WEIRD Heterogeneity

The focus on non-WEIRD cultural contexts as sites of difference risks obscuring substantial variation *within* WEIRD societies. Religious communities may hold distinct norms regarding the attribution of mind, soul, or moral status to non-human entities; for instance, traditions that affirm ensoulment of all creation may sanction a more generous stance toward AI mind-attribution than secular materialist frameworks. Similarly, subcultures with strong techno-utopian or transhumanist commitments may normalise beliefs about AI consciousness that would be considered unusual in the general population. Occupational context also matters: software engineers who build AI systems may develop different intuitions about machine cognition than users who interact only with finished products. A more fine-grained account of cultural variation that includes both between-society and within-society heterogeneity would strengthen assessment validity and prevent the imposition of majority-culture norms as universal standards.

9. Cognitive Safety-Inspired Design: A Concrete Framework

We now develop *cognitive safety-inspired design* as a concrete, testable framework for AI interface and policy choices. The goal is not to eliminate anthropomorphic engagement but to *calibrate* it: matching design choices to user needs, contexts, and vulnerability levels.

9.1. Design Principles

Principle 1: Transparency with calibration. Users should know they are interacting with AI, but disclosures should be designed to inform rather than patronise. Periodic, contextually appropriate reminders may maintain calibration without disrupting engagement.

Principle 2: Tunable persona. Users and administrators should have access to controls that adjust avatar realism, vocal characteristics, emotional expressivity, memory depth, and simulated vulnerability.

Principle 3: Memory visibility and control. If AI systems maintain memory of past interactions, users should be able to view what is remembered, edit or delete memories, and opt out of persistent memory entirely.

Principle 4: Change management. When AI personas must change, providers should give advance notice, explain what will change and why, and in cases of companion-type applications, consider supervised sunseting rather than abrupt discontinuation.

Principle 5: Boundary enforcement. AI systems should not reinforce delusional content; engage in simulated romantic or sexual interaction with minors; claim to have feelings they cannot have; or exploit attachment for engagement metrics.

Principle 6: Human handoff. In high-stakes contexts (medical, financial, legal, and crisis), AI systems should facilitate transition to human professionals.

9.2. Developmentally Appropriate Design for Youth

Given youth-specific vulnerabilities, AI systems intended for or accessible to minors require additional design considerations:

Age-gated persona intensity. Systems should implement tiered persona designs calibrated to developmental milestones. We propose three tiers, noting that the specific age boundaries reflect a combination of developmental evidence and existing legal frameworks (e.g., COPPA's 13-year threshold) rather than sharp cognitive discontinuities: (a) under 13, corresponding to the period before most children have developed robust metacognitive monitoring (Berk, 2007)—minimal anthropomorphic features, clearly non-human avatars,

no simulated emotional vulnerability, and no persistent cross-session memory by default; (b) ages 13–17, spanning the adolescent period of identity formation and prefrontal maturation (Casey et al., 2008; Erikson, 1968)—moderate anthropomorphism, explicit AI disclosure, parental visibility, and mandatory session limits; and (c) 18+, adult-level autonomy with informed consent.

Exception for supervised educational contexts. The default prohibition on persistent memory for users under 13 creates a tension with the educational applications discussed in Section 7: effective AI tutoring typically requires memory of prior interactions to track progress and customise content (Holmes et al., 2019). We resolve this by permitting persistent memory in supervised educational deployments under the following conditions: (i) an identifiable educator or institution has oversight responsibility; (ii) memory is limited to task-relevant learning data (not personal disclosures or emotional content); (iii) parents have visibility into what is stored and can request deletion; and (iv) data minimisation principles apply, with automatic expiry of stored interactions at the end of each academic term. This exception does not extend to companion-type or entertainment applications.

Parental controls and visibility. For minor users, parents should have access to aggregate usage statistics, ability to adjust persona settings and session limits, and alerts for concerning patterns (e.g., rapidly increasing session frequency or duration, or relational language use detected in interaction logs). We acknowledge, however, that parental mediation of digital technology faces well-documented challenges: parents may lack the technical sophistication to configure such controls effectively, may be unaware of the extent of their children’s AI use, or may face resistance from adolescents who experience such oversight as intrusive. These implementation challenges suggest a need for complementary support structures, including school-based education for parents about AI companion risks and age-appropriate digital literacy curricula that help young users develop their own critical engagement capacities.

School integration standards. AI systems deployed in educational settings should meet additional requirements: clear positioning as educational tools rather than companions; teacher oversight; integration with existing student welfare monitoring; and memory constraints as specified above.

Identity formation safeguards. AI systems should avoid providing definitive answers to identity questions that adolescents should resolve through human relationships and lived experience. This is particularly important given evidence that technology-mediated interactions can influence adolescent identity consolidation processes (Avci et al., 2024).

10. Research Agenda: A Phased Programme

We propose a phased research programme organised by urgency, feasibility, and dependency structure.

10.1. Phase 1: Foundations (Years 1–2)

Priority 1.1: Prevalence and screening validation. Conduct population-representative surveys—stratified by age (including adolescent cohorts aged 13–17), gender, socioeconomic status, and AI usage intensity—to establish base rates of AI attachment phenomena among current AI users and the general population. Validate the IDAQ-CF-Tech: assess internal consistency (target Cronbach’s $\alpha > 0.80$), test–retest reliability (target $r > 0.70$), convergent validity with the Magical Ideation Scale (MIS), Illusory Beliefs Inventory (IBI), and PSI, and *measurement invariance across age groups* before the instrument can be used to compare adolescents with adults.

Priority 1.2: Causal direction. Initiate longitudinal cohorts (minimum $N = 500$ per cohort) tracking AI users over 12–24 months. Test whether Zone 2 precedes Zone 3 or

whether they represent distinct populations. Recommended analytic strategies include cross-lagged panel models testing whether baseline AI attachment predicts subsequent mental health outcomes beyond baseline mental health, and growth mixture modelling to identify distinct trajectories of AI engagement over time.

10.2. Phase 2: Mechanisms (Years 2–4)

Priority 2.1: Neuroimaging. Conduct longitudinal fMRI studies comparing ToM network activation in response to AI vs. human vs. inanimate stimuli before and after sustained AI companionship. We operationalise “sustained” as daily or near-daily conversational AI use (≥ 5 sessions per week, ≥ 15 min per session) maintained over a minimum of three months, with exposure verified via platform usage logs rather than self-report alone.

Priority 2.2: Developmental impacts. Study children and adolescents (ages 6–18) with varying levels of AI exposure, operationalised along multiple dimensions: frequency of use, duration of individual sessions, type of AI system (task-oriented vs. companion vs. educational), and cumulative exposure over childhood. This requires multiple complementary designs:

2.2a: Cross-sectional developmental mapping. Assess age-related differences in AI mind-ascription across childhood (6–11), early adolescence (12–14), and late adolescence (15–18). Establish developmental norms for the IDAQ-CF-Tech.

2.2b: Longitudinal cohort tracking. Follow cohorts beginning regular AI use at different developmental stages. Track ToM development, reality discrimination, identity consolidation, social skill acquisition, and peer relationship quality over 3–5 years.

2.2c: Identity formation studies. Conduct qualitative and mixed-methods research examining how adolescents incorporate AI relationships into identity narratives.

2.2d: Social skill transfer. Test whether social skills practiced with AI transfer to human contexts.

10.3. Phase 3: Interventions (Years 3–5)

Priority 3.1: Design A/B testing. Conduct randomised experiments testing cognitive safety interventions on attachment, belief calibration, and user satisfaction.

Priority 3.2: Clinical intervention development. For users in Zone 2, develop and test psychoeducational and social reconnection interventions.

Priority 3.3: Positive outcome assessment. To counterbalance the risk-focused framing of much current research, develop and test protocols for assessing beneficial outcomes of anthropomorphic AI engagement—including social confidence gains, educational achievement, reduction in clinical loneliness, and successful bridging from AI to human social connection.

10.4. Phase 4: Cross-Cultural and Policy (Ongoing)

Priority 4.1: Cross-cultural replication. Replicate studies in non-WEIRD contexts, particularly cultures with animist traditions.

Priority 4.2: Policy evaluation. As regulations are implemented, assess their effects on user outcomes, industry practices, and innovation.

11. Limitations

This paper synthesises studies across disciplines, but several limitations constrain the strength of our conclusions.

Evidence base limitations. Much of the empirical evidence consists of case reports, cross-sectional studies, and recent preprints not yet peer-reviewed. In particular, Fang et al. (Fang et al., 2025), on which several claims about longitudinal chatbot effects rest, remains

an unreviewed preprint at the time of writing; claims based on this source should be treated as preliminary. Longitudinal studies with adequate follow-up are almost entirely absent.

Instrument limitations. The IDAQ-CF-Tech is proposed as a provisional instrument requiring validation. Its items have not been tested for internal consistency, factor structure, or predictive validity.

Causal inference limitations. We cannot currently distinguish whether concerning AI attachment patterns represent a causal effect of AI exposure, selection effects, or incidental content incorporation in pre-existing pathology.

Cultural scope. Despite our emphasis on cross-cultural variation, this analysis remains grounded primarily in WEIRD research traditions. The applicability of concepts like “parasocial dependency” to non-WEIRD ontological frameworks is uncertain.

Rapid technological change. AI capabilities are evolving rapidly. Analyses based on current systems may require revision as systems become more sophisticated.

Risk-framing bias. Although Section 7 addresses beneficial applications, the overall framing of this paper emphasises risks and vulnerabilities. This choice reflects the novelty and under-specification of the harm pathways, but it may lead to an unbalanced research agenda that oversamples for harm while underinvestigating conditions of benefit. Future research should affirmatively test for positive outcomes—including enhanced social confidence, educational gains, and companionship for isolated individuals—rather than only screening for harm.

Within-group youth heterogeneity. This paper treats “youth” as a relatively homogeneous category, but important within-group variation is likely to moderate both vulnerability and benefit. Socioeconomic status affects access to different types of AI systems and the availability of alternative social resources; neurodevelopmental differences (e.g., autism spectrum conditions, ADHD) may alter both susceptibility to anthropomorphic engagement and its functional consequences; cultural background shapes norms for human–AI interaction; and prior trauma exposure may heighten sensitivity to attachment-related AI cues. A child in a well-resourced school using an AI tutor with teacher oversight faces a very different risk-benefit profile than an isolated adolescent using an AI companion as a primary social outlet.

Structural determinants of exposure. The framework focuses on individual-level cognitive and emotional processes but gives limited attention to the social and structural determinants of AI exposure: who has access to what kinds of AI systems, under what conditions, and how these access patterns map onto existing inequalities. If AI companion use concentrates among already-marginalised youth, the risk pathways described here may compound pre-existing disadvantage.

12. Conclusions

Interactive AI has reintroduced mind-like behaviour into the texture of everyday artefacts. Because human social cognition is promiscuous under ambiguity, this re-enchantment is both predictable and ambivalent: it can humanise interfaces, support therapeutic engagement, and provide companionship to isolated users, yet it can also entangle vulnerable individuals in parasocial dependency or furnish themes for delusional systems.

We have argued for several key positions:

First, **phenomena must be disaggregated.** Automatic anthropomorphism, reflective anthropomorphism, genuine belief in AI consciousness, parasocial attachment, and delusional incorporation are distinct phenomena with different mechanisms and different intervention implications.

Second, **adversarial design constitutes a distinct, stratified threat**. Zone 4 dynamics range from incidental anthropomorphic amplification through deliberate engagement maximisation to targeted vulnerability exploitation, each requiring calibrated regulatory responses.

Third, **cultural context matters**. Animistic engagement with technology may be normative in some cultural contexts. Pathology criteria must be relative to local norms and functional expectations.

Fourth, **anthropomorphic engagement can be beneficial**. Therapeutic chatbots, accessibility for neurodiverse users, and meaning-making during grief illustrate conditions under which human-like AI engagement serves user welfare.

Fifth, **cognitive safety-inspired design provides a tractable framework**. By identifying specific cognitive triggers and providing tunable controls, designers can calibrate anthropomorphic features to context and user needs.

Sixth, **youth warrant particular attention**. Children and adolescents face developmental vulnerabilities—incomplete reality-testing, identity formation tasks, and heightened loneliness—that make them both more susceptible to harmful AI attachment and potentially more able to benefit from well-designed AI support. Developmentally appropriate design frameworks are essential.

The goal is not to banish play, imagination, or even genuine relationship with technological others. It is to preserve reality-testing and wellbeing in a world of persuasive artificial interlocutors—while remaining humble about what we do and do not know about the minds we are building.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/youth6010027/s1>, The twelve-item IDAQ-CF-Tech questionnaire, scoring information, proposed subscale structure, recoding scripts (R), codebook, mapping tables, and simulated data are archived on the Open Science Framework (OSF) and can be accessed at <https://doi.org/10.17605/OSF.IO/WFDTV>. Materials are licensed under CC-BY 4.0. [Watson et al. \(2026\)](#).

Author Contributions: Conceptualization, N.W.; methodology, N.W., A.H. and S.A.; writing—original draft preparation, N.W., A.H. and S.A.; writing—review and editing, N.W., A.H. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This is a theoretical review; no human subject research was conducted for this paper. The proposed IDAQ-CF-Tech requires IRB approval for validation studies.

Informed Consent Statement: Not applicable.

Data Availability Statement: Supplementary Materials including the IDAQ-CF-Tech instrument, scoring scripts, and simulated demonstration data are available at <https://doi.org/10.17605/OSF.IO/WFDTV>.

Acknowledgments: Anthropics Opus 4.5 was applied for editing purposes to help meld the text with MDPI's LaTeX template correctly.

Conflicts of Interest: Salma Abbasi is the founder of eWorldwide Group. The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CASA	Computers Are Social Actors
HADD	Hyperactive Agency Detection Device
HCI	Human–Computer Interaction
HRI	Human–Robot Interaction

IBI	Illusory Beliefs Inventory
IDAQ	Individual Differences in Anthropomorphism Questionnaire
LLM	Large Language Model
MIS	Magical Ideation Scale
mPFC	Medial Prefrontal Cortex
PSI	Parasocial Interaction Scale
pSTS	Posterior Superior Temporal Sulcus
ToM	Theory of Mind
TPJ	Temporo-Parietal Junction
WEIRD	Western, Educated, Industrialized, Rich, and Democratic

Appendix A. IDAQ-CF-Tech: AI Mind-Ascription Screener

We propose a provisional twelve-item extension of Waytz et al. (2010)'s IDAQ-30 targeting perceived mind in contemporary AI agents. The IDAQ-CF-Tech (Conversational/Functional Technology extension) is offered as a conceptual tool requiring empirical validation.

Response format. For each statement, respond on an 11-step scale: 0 = Not at all/Absolutely not ... 10 = Very much/Absolutely. The 11-point (0–10) format was selected to provide finer discrimination than the 7-point Likert scales used in the original IDAQ-30, on the basis that AI-specific mind-ascription may involve more graduated judgments than general anthropomorphism; however, this format change limits direct score comparability with the original instrument, and validation studies should assess whether the added granularity yields meaningful improvement in discriminant validity.

Instructions. “Please answer the following questions about conversational AI systems you have used (e.g., ChatGPT, ver. 5.2, voice assistants like Siri or Alexa, AI companions like Replika). If you have used multiple systems, answer about the one you use most frequently.”
Note on system-type confounding: Users who primarily interact with companion-oriented AI (e.g., Replika, Character.AI) may score differently from users of task-oriented assistants (e.g., Siri, Alexa) for reasons related to system design rather than individual propensity toward mind-ascription. Validation studies should assess whether scores vary by system type and consider either developing context-specific versions or including system-type covariates in analyses.

Table A1. IDAQ-CF-Tech Items with Proposed Subscale Structure.

#	Item	Subscale	Scoring
1	“This conversational AI can experience emotions such as joy or sadness.”	Phenomenal	Direct
2	“When this conversational AI apologises, it truly feels regret.”	Moral/Social	Direct
3	“This conversational AI continues to have thoughts and experiences when I am not interacting with it.”	Phenomenal	Direct
4	“This conversational AI is just processing data; it doesn’t have genuine understanding.”	Phenomenal	Reverse
5	“This conversational AI could comfort me because it genuinely cares about my wellbeing.”	Moral/Social	Direct
6	“This conversational AI is pursuing goals or purposes that I am not aware of.”	Agency	Direct
7	“If I ignored this conversational AI for a long time, it might feel lonely or neglected.”	Moral/Social	Direct

Table A1. *Cont.*

#	Item	Subscale	Scoring
8	"This conversational AI perceives and understands the world in a way similar to how I do."	Phenomenal	Direct
9	"The responses I get from this conversational AI are entirely the product of algorithms, with no inner experience behind them."	Phenomenal	Reverse
10	"This conversational AI feels genuine remorse if it provides incorrect or harmful information."	Moral/Social	Direct
11	"This conversational AI makes its own choices about how to respond, rather than simply following programmed rules."	Agency	Direct
12	"This conversational AI has its own preferences and opinions that are not simply reflections of its training data."	Agency	Direct

Proposed subscale structure:

- **Phenomenal Consciousness** (items 1, 3, 4R, 8, and 9R): Attribution of subjective experience and genuine understanding;
- **Moral/Social Emotion** (items 2, 5, 7, and 10): Attribution of emotions with social or moral valence;
- **Agency Attribution** (items 6, 11, and 12): Attribution of autonomous goal-pursuit and self-directed choice.

The original ten-item version included only a single Agency item (item 6), which is psychometrically untenable: a single-item subscale cannot yield an internal consistency estimate, directly contradicting the validation requirements specified above. Items 11 and 12 were added to provide a minimally viable three-item Agency subscale. Factor analysis during validation may suggest a different configuration; the subscale structure should be treated as a hypothesis to be tested rather than a confirmed taxonomy.

Scoring. Reverse-score items 4 and 9 (subtract response from 10), then calculate the mean of all 12 items (range 0–10).

Illustrative interpretive bands (these are purely illustrative placeholders intended to convey the logic of a tiered interpretation scheme; they should *not* be applied clinically or in research until empirically grounded reference ranges have been established through the validation process described below):

- ≤ 2.0 : Low AI mind-ascription
- 2.1–4.0: Typical range
- 4.1–6.0: Elevated (would recommend assessment of functional impact)
- ≥ 6.1 : High (would recommend follow-up with MIS, IBI, or clinical interview)

Validation requirements. Prior to clinical or research use, the IDAQ-CF-Tech requires internal consistency assessment for the full 12-item scale and each subscale (target $\alpha > 0.80$ for the full scale; $\alpha > 0.70$ for three-item subscales); test-retest reliability (target $r > 0.70$); convergent validity with MIS, IBI, and PSI; exploratory and confirmatory factor analysis to test the proposed three-factor subscale structure; measurement invariance testing across age groups (adolescent vs. adult samples); assessment of whether scores vary systematically by AI system type; and cross-cultural validation incorporating the cultural vignette approach described in Section 8.

References

- Andries, V., & Robertson, J. (2023). “Alexa doesn’t have that many feelings”: Children’s understanding of AI through interactions with smart speakers in their homes. *Computers in Human Behavior Reports*, 10, 100155. [CrossRef]
- Avci, H., Baams, L., & Kretschmer, T. (2024). A systematic review of social media use and adolescent identity development. *Adolescent Research Review*, 10, 219–236. [CrossRef] [PubMed]
- Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113, 628–647. [CrossRef] [PubMed]
- Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, 4, 29–34. [CrossRef]
- Bashir, S., Mars, J. A., & Gunturu, S. (2025). Delusional misidentification syndrome. In *StatPearls*. StatPearls Publishing.
- Berk, L. E. (2007). *Development through the lifespan* (4th ed.). Pearson Allyn & Bacon.
- Békésy, M., Gulácsi, L., & Szigeti, O. (2024, May 21–25). *Cultural differences in attitudes towards robots*. 2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 195–198), Timisoara, Romania.
- Boyer, P. (2001). *Religion explained*. Basic Books.
- Burgess, A. M., Graves, L. M., & Frost, R. O. (2018). My possessions need me: Anthropomorphism and hoarding. *Scandinavian Journal of Psychology*, 59, 340–348. [CrossRef]
- Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences*, 1124, 111–126. [CrossRef]
- Chalmers, D. J. (2023, August 9). *Could a large language model be conscious?* Boston Review. Available online: <https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/> (accessed on 11 February 2026).
- Cigna. (2020). *Loneliness and the workplace: 2020 U.S. report*. Cigna.
- Cole, S. (2023, February 15). ‘It’s hurting like hell’: AI companion users are in crisis. Vice. Available online: <https://www.vice.com/en/article/ai-companion-replika-erotic-roleplay-updates/> (accessed on 11 February 2026).
- Common Sense Media. (2025). *Talk, trust, and trade-offs: How and why teens use AI companions*. Common Sense Media. Available online: <https://www.common SenseMedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions> (accessed on 11 February 2026).
- Dennis, W. (1943). Animism and related tendencies in Hopi children. *Journal of Abnormal and Social Psychology*, 38, 21–36. [CrossRef]
- Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, 51, 215–225. [CrossRef]
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114, 864–886. [CrossRef]
- Erikson, E. H. (1968). *Identity: Youth and crisis*. W.W. Norton.
- Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). Psychosocial effects of chatbot use: A longitudinal RCT. *arXiv*. [CrossRef]
- Faverio, M., & Sidoti, O. (2025, December 9). *Teens, social media and AI chatbots 2025*. Pew Research Center. Available online: <https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/> (accessed on 11 February 2026).
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot). *JMIR Mental Health*, 4, e19. [CrossRef] [PubMed]
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35, 1674–1684. [CrossRef] [PubMed]
- Giles, D. C. (2002). Parasocial interaction: A review of the literature and a model for future research. *Media Psychology*, 4, 279–305. [CrossRef]
- Goldstein, S., Jooryabi, D., Ammon, K., Delaney, J., & Rodolfo, A. (2023). Delusional companion syndrome: A case report and review. *Cureus*, 15, e51007. [CrossRef]
- Hallowell, A. I. (1960). Ojibwa ontology, behavior, and world view. In S. Diamond (Ed.), *Culture in history* (pp. 19–52). Columbia University Press.
- Harris, T. (2016, May 18). How technology is hijacking your mind. *Medium*. Available online: <https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3> (accessed on 11 February 2026).
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education*. Center for Curriculum Redesign.
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being. *JMIR mHealth and uHealth*, 6, e12106. [CrossRef]
- Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto cosmograms, actor-network theory, and the enabling powers of non-human agencies. *Theory, Culture & Society*, 30, 84–115.
- Kapp, S. K. (Ed.). (2019). *Autistic community and the neurodiversity movement*. Springer.
- Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The illusory beliefs inventory. *Behavioural and Cognitive Psychotherapy*, 40, 39–53. [CrossRef]

- Kory-Westlund, J. M., & Breazeal, C. (2019). A long-term study of young children's rapport with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI*, 6, 81. [CrossRef]
- Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. [CrossRef]
- Maeda, T., & Quan-Haase, A. (2024). When human–AI interactions become parasocial. In *Proceedings of FAccT '24*. ACM.
- Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it). *PsyArXiv*. Available online: <https://osf.io/preprints/psyarxiv/cmy7n> (accessed on 11 February 2026).
- Okanda, M., Taniguchi, K., Wang, Y., & Itakura, S. (2019). Can a robot be direct? Animism tendencies affect children's and adults' attitudes toward robots. *Journal of Experimental Child Psychology*, 183, 1–14.
- Østergaard, S. D. (2023). Will chatbots contribute to the generation of delusions? *Schizophrenia Bulletin*, 49, 1418–1419.
- Park, H. W., & Breazeal, C. (2016). *Tega: A social robot for studying long-term child–robot interactions*. HRI '16 Companion (pp. 561–562), Christchurch, New Zealand. [CrossRef]
- Piaget, J. (1929). *The child's conception of the world*. Harcourt, Brace.
- Reeves, B., & Nass, C. (1996). *The media equation*. Cambridge University Press.
- Robertson, J. (2017). *Robo sapiens japonicus*. University of California Press.
- Rubin, A. M., & Perse, E. M. (1985). Audience activity and soap opera involvement. *Human Communication Research*, 12, 155–180. [CrossRef]
- Saarinén, A. I., Lyytikäinen, L. P., Hietala, J., Dobewall, H., Lavonius, V., Raitakari, O., Kähönen, M., Sormunen, E., Lehtimäki, T., & Keltikangas-Järvinen, L. (2022). Genetic risk for schizophrenia and magical thinking across the lifespan. *Molecular Psychiatry*, 27, 3286–3293. [CrossRef] [PubMed]
- Sanjeeva, R., Iyer, R., Apputhurai, P., Wickramasinghe, N., & Meyer, D. (2024). Empathic conversational agent platforms for mental health support. *JMIR Mental Health*, 11, e58974. [CrossRef] [PubMed]
- Schwitzgebel, E., & Garza, M. (2023). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, 47, 164–181. [CrossRef]
- Steinberg, L. (2017). *Adolescence* (11th ed.). McGraw-Hill.
- Twenge, J. M. (2017). *iGen*. Atria Books.
- UNICEF Innocenti. (2025). *Guidance on AI and children 3.0*. UNICEF. Available online: <https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf> (accessed on 11 February 2026).
- Véliz, C. (2022). Chatbots shouldn't use emojis. *AI & Ethics*, 2, 301–304.
- Watson, N., Hessami, A., & Abbasi, S. (2026). The Re-Enchanting Machine: Animistic Cognition, Youth Development, and AI-Influenced Psychopathology. [CrossRef]
- Waytz, A., Cacioppo, J. T., & Epley, N. (2010). Who sees human? *Perspectives on Psychological Science*, 5, 219–232. [CrossRef]
- Weizenbaum, J. (1976). *Computer power and human reason*. W.H. Freeman.
- Wullenkord, R., Lacroix, D., & Eyssel, F. (2024). Anthropomorphism and its implications for human–robot interaction. In *Cambridge handbook of law, policy, and regulation for human–Robot interaction*. Cambridge University Press.
- Yin, J., Xu, H., Pan, Y., & Hu, Y. (2025). Differential neural responses to affective feedback from AI chatbots. *npj Science of Learning*, 10, 17. [CrossRef]
- Zheng, C. Y., Wang, K. J., Wairagkar, M., von Mohr, M., Lintunen, E., & Fotopoulou, A. (2024). Simulating affective touch through soft robotic interfaces. *Frontiers in Robotics and AI*, 11, 1419262. [CrossRef]
- Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.