Social Science Chronicle

Peer-Reviewed, International, Academic Research Journal

ISSN : 3048-6297

RESEARCH ARTICLE

Human-Centered Artificial Intelligence: Integrating Human Cognition, Psychological Assessment and Responsible Digital Behavior for Real World Well-Being

Ramya Ali Pashtun[1]* and Nadeesha Nadarajah[2]

[1] University of Kashmir, Srinagar, India.
[2] Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.

* *Corresponding Author*
E-mail : ramyapashtun@hotmail.com

Citation
Pashtun, R. A., & Nadarajah, N. (2026). Human-Centered Artificial Intelligence: Integrating Human Cognition, Psychological Assessment and Responsible Digital Behavior for Real World Well-Being. *Social Science Chronicle*, Vol. 6, Issue - 1, pp. 1-12.

Digital Object Identifier (DOI)
https://doi.org/10.56106/ssc.2026.005

Received - October 25, 2025
Accepted - February 14, 2026
Published - Online First

Web-Link
All contents of this peer-reviewed article, along with other relevant details are available at:
http://socialsciencechronicle.com/article-ssc-2026-005

Abstract

This article enunciates a construct-centered synthesis of *artificial intelligence* as a psycho-technical infrastructure that increasingly mediates cognition, affect, motivation, identity, and social coordination across clinical, educational, organizational, and civic ecologies. It argues that psychologically consequential AI must be governed first as a *construct validity* problem and only second as a predictive optimization problem, because statistically impressive models can remain ontologically mis-specified when they operationalize socially contingent proxies, ignore contextual meaning, or induce feedback-driven performativity. The article builds an epistemic bridge between psychological explanation and algorithmic prediction by translating *nomological networks*, reliability, *measurement invariance*, and counterfactual reasoning into design-grade requirements for AI systems that assess, profile, recommend, or intervene. It then reframes AI as an implicit theory-engine of mind and behavior, integrating *dual-process* cognition, *signal detection*, evidence accumulation, reinforcement dynamics, and *extended cognition* to explain predictable human-AI failure modes such as automation bias, miscalibrated trust, attentional capture, and dependency formation. The article further specifies how psychometrics, affective inference, and trait profiling can fail under proxy confounding, dataset shift, concept drift, and explanation illusions, and it delineates mechanism-fidelity constraints for AI-mediated intervention grounded in process-based change, autonomy support, and iatrogenic risk minimization. Finally, it advances a global governance lens that treats psychological integrity, equity, contestability, and lifecycle auditability as enforceable socio-technical obligations, offering a cross-sector roadmap for academically rigorous, policy-relevant, and implementation-ready psychologically safe AI.

Keywords
Artificial Intelligence, Affective Computing, Human-AI Interaction, Psychological Assessment, Psychometrics, Digital Phenotyping, Emotion Regulation, Cognitive Modelling, Reinforcement Learning, Algorithmic Fairness, Trust Calibration, Mental Health Technology.

1. Introduction

Contemporary *artificial intelligence* has moved from peripheral automation to continuous, ambient mediation of attention, choice, emotion, and identity, which makes *psychological theory* a design constraint rather than an interpretive afterthought. This shift is occurring against a global mental health burden that is no longer plausibly treated as a niche clinical concern. More than one billion people are living with mental health disorders worldwide, depression affects about 332 million people, anxiety disorders are estimated at about 359 million people, and suicide deaths in 2021 were estimated at 727,000, all of which makes prevention, triage, and support infrastructures a policy-level necessity. Digital mental health ecosystems have also scaled rapidly, with over 10,000 mental health and wellness apps circulating in largely heterogeneous quality regimes, intensifying the need for *construct-valid* psychological measurement and *harm-aware* intervention logic. Simultaneously, organizational adoption is rising, with OECD-reported AI use among firms with 10 or more employees increasing from 5.6 percent in 2020 to 14 percent in 2024, creating ubiquitous workplace psychodynamics around *trust-calibration*, *surveillance stress* and *algorithmic authority*.

An article on AI and psychology must be written for multiple epistemic communities whose incentive structures differ, yet whose decisions converge on human outcomes. For academics and researchers, the core deliverable is a conceptual synthesis that links *nomological networks* to AI system properties, clarifying where *predictive validity* diverges from *explanatory coherence*, and why psychological inference requires more than error minimization (Piotrowski, 2025; Paesano, 2023; Priya & Sharma, 2023). For industry experts, the deliverable is an actionable psychological language for product architecture, centered on *human factors engineering*, *cognitive ergonomics*, *behavioral design*, and *risk-controls* that prevent foreseeable failures such as *automation bias*, *reactance*, and *dependence formation*. For policymakers, the deliverable is a governance-ready taxonomy of harms that treats *psychological integrity*, *autonomy*, and *equity* as enforceable requirements rather than aspirational ethics. For students and workforce development professionals, the deliverable is an integrative scaffold that connects *cognitive science*, *affective science*, *social psychology*, *clinical models*, and *policy analytics* into a coherent interdisciplinary grammar, enabling competent participation in high-stakes design, procurement, regulation, and evaluation.

Central Argument, Organizing Questions, and the Primacy of Construct Validity

The central argument is that AI in psychologically sensitive domains is first a *construct validity* problem and only then a performance optimization problem, because systems that predict well can still be psychologically wrong, unjust, or unsafe if they operationalize the wrong construct, compress context into misleading proxies, or induce adverse feedback loops. This reframes evaluation from an accuracy-only mindset toward a triad of *validity*, *safety*, and *institutional accountability*. The organizing questions are designed to be decision-relevant across sectors. What psychological constructs are being inferred, and what alternative constructs could explain the same digital signal. When is prediction ethically acceptable without mechanistic explanation, and when is explanation a moral requirement because the system's output becomes a justificatory artifact in high-stakes decisions. How should uncertainty be represented to match human *metacognitive limits* and reduce *overreliance*. Which design choices create *performativity*, where an AI label reshapes self-perception and behavior until the label appears confirmed. These questions also foreground *error-cost asymmetry*, since psychological harms from false reassurance, stigma amplification, or coerced disclosure can be qualitatively non-reversible.

Core Definitions, Boundary Conditions, and Two-Level Analytical Discipline

This article treats *AI* as a family of socio-technical systems that includes predictive models, generative models, conversational agents, and adaptive recommender architectures, each with different psychological affordances and failure modes. The psychological scope spans intra-individual processes such as *attention*, *working memory*, *emotion regulation*, and *self-control*, interpersonal processes such as *attribution*, *trust*, and *social influence*, and macro-level processes such as *norm formation*, *institutional power*, and *cultural meaning systems*. The boundary condition is conceptual-theoretical rather than empirical, meaning the article prioritizes definitional precision, mechanistic plausibility, and governance logic over study-by-study reporting. It also disciplines the analytic hierarchy to two levels only, ensuring every subsection functions as a self-contained conceptual unit with a single argumentative arc. Within this constraint, the article treats *psychometrics*, *causal inference*, and *human factors* as foundational languages, because they translate abstract psychological constructs into operational requirements for measurement, intervention, and accountability without reducing psychology to instrumentation.

Roadmap and Integrative Throughline

The article proceeds through a tightly coupled sequence that preserves conceptual continuity while spanning disciplines. Section 2 establishes epistemic bridgework, aligning psychological explanation, measurement theory, and causal reasoning with AI prediction, uncertainty, and socio-technical deployment. Section 3 reframes AI as a partial model of mind and social cognition, mapping *dual-process architectures*, *Bayesian cognition*, *reinforcement learning*, and *extended cognition* to real-world human-AI interaction dynamics. Section 4 develops a construct-centered account of assessment, emphasizing *reliability*, *measurement invariance*, *proxy confounding*, and the distinctive validity threats of affective and personality inference. Section 5 translates psychotherapy and behavior change theories into design constraints for AI-mediated intervention, foregrounding *mechanisms of change*, *autonomy support*, and *iatrogenic risk*. Section 6 scales outward to societal, organizational, and policy dimensions, treating equity, privacy, psychological safety, and governance as determinants of downstream impact rather than externalities. Section 7 synthesizes a unified set of principles that are simultaneously theoretical, implementable, and auditable across global contexts.

2. Epistemic Foundations and Conceptual Bridgework between AI and Psychology

Psychological explanation privileges *mechanism*, *process*, and *counterfactual dependence*, whereas many AI pipelines privilege predictive optimization under a loss function, creating an epistemic asymmetry that becomes consequential in psychologically sensitive decisions. A system can be accurate and well-calibrated yet ontologically mis-specified, because its features may encode platform affordances, institutional incentives, or discourse conventions rather than the intended latent construct (Matz et al., 2024; Al-Mseidin et al., 2023). This reframes model evaluation as a problem of *abductive inference* under severe underdetermination, where multiple psychological stories can fit the same statistical surface. In such settings, prediction should be treated as instrumental, not evidentiary, unless it is tethered to theory-consistent constraints on what counts as a plausible proxy. Calibration, discrimination, and stability are necessary but insufficient when proxies are socially contingent. A disciplined bridge therefore requires explicit claims about the target construct, admissible indicators, and boundary conditions for generalization, including construct drift, distributional shift, strategic behavior, and endogenous feedback from users and institutions.

Construct Validity, Psychometric Reliability, and the Nomological Network Translation Layer

The central translation layer between AI outputs and psychological meaning is *construct validity*, articulated through a coherent *nomological network* that specifies lawful relations among constructs, indicators, and outcomes. Measurement theory treats scores as products of an instrument-plus-context process with quantifiable error, so AI-based inference must engineer *reliability* across time, devices, languages, and interaction settings (Harding et al., 2024; Schwesig et al., 2023). *Convergent validity* and *discriminant validity* become multimodal constraints, requiring that textual, behavioral, and physiological signals converge only when theory predicts shared variance, and remain separable when constructs are adjacent but distinct. *Content validity* and *criterion validity* then constrain feature selection, outcome alignment, and error-cost asymmetry in high-stakes screening. *Measurement invariance* and *differential item functioning* generalize into feature-space audits, ensuring that the same psychological score does not mean different things across cultures, neurotypes, or socioeconomic contexts. Psychometric logics from *classical test theory* and *item response theory* thus operate as governance primitives for AI claims about emotion, personality, and risk.

Psychological phenomena are irreducibly multilevel, spanning intra-individual micro-dynamics, dyadic interaction, group norms, organizational climates, and cultural meaning systems, while AI models often compress these strata into a single feature space, inviting cross-level fallacies. Individual predictions can be illegitimately scaled into population policy, and population priors can be illegitimately imposed on individuals, producing epistemic error and distributive injustice (Presbitero & Teng-Calleja, 2023; Marx et al., 2023). The person-situation problem is reactivated under algorithmic mediation, because platform design can increase *situational strength* and suppress behavioral variability that would otherwise reveal stable dispositions. *Ecological validity* becomes a binding constraint, since digital traces are co-produced by interface architecture, incentive gradients, and attention economies, not merely by persons. A theoretically faithful bridge therefore requires explicit context models, a social-ecological lens for multilevel causation, attention to institutional logics, recognition of emergent norm formation, and safeguards against context collapse, including role-based interpretation, temporal granularity, and culturally grounded semantics for what behaviors signify.

Causal Inference, Counterfactual Semantics, and Intervention-Ready Psychological Knowledge

When AI recommends, nudges, triages, or treats, the relevant epistemic object is intervention, not prediction, which requires *causal inference* rather than correlational patterning. Psychological theory already separates mediators from moderators, and this distinction should be treated as an engineering specification, because mediators encode mechanisms of change, while moderators encode effect heterogeneity across persons and contexts. Counterfactual reasoning grounded in the *potential outcomes* framework clarifies that a model output is actionable only if it supports credible claims about what would happen under alternative choices. Conceptual tools such as *directed acyclic graphs*, confounding control, collider awareness, and, where conceptually defensible, instrumental leverage, become necessary to avoid policy-meaningless correlations, especially in platforms with adaptive personalization and reciprocal causation (Gupta & Guglani, 2023; Tassiello et al., 2025). Dynamic psychological systems further demand attention to time, dosage, and sequencing, motivating *dynamic treatment regimes* and mechanism-aware monitoring that anticipates feedback loops, performativity, and unintended reinforcement of avoidance, dependence, or stigma.

Human Factors, Explanation Psychology, and Normative Governance as Psychological Safety Infrastructure

Even a technically competent model can fail psychologically through predictable human factors dynamics, including *automation bias*, vigilance decrement, authority gradients, reactance, and miscalibrated trust, so cognition-aware interface design is part of epistemic validity. Users evaluate explanations through coherence heuristics, cognitive load limits, and plausibility filters, which can generate an illusion of understanding and unjustified compliance, particularly when outputs are framed with spurious precision (Watermann et al., 2025; Samoilenko & Suvorova, 2023). The bridge therefore requires *trust calibration* mechanisms, uncertainty communication, contestability pathways, and workload-aware workflows that reduce error cascades and clinician moral injury in delegated decision chains. Normative governance must translate ethical principles into operational requirements, treating autonomy as *self-determination* and nonmaleficence as harm minimization across iatrogenic risks such as labeling effects, identity threat, and surveillance-induced anxiety. Procedural justice further demands explainable decision rights, auditable risk registers, and accountability allocation across designers, deployers, and institutions, so responsibility does not diffuse into a sociotechnical fog.

3. AI as a Theory-Engine and Model of Mind, Behavior, and Social Cognition

A psychologically literate AI architecture is not merely a statistical function approximator, it is an implicit model of cognition that can be interrogated through *computational rationality* and *cognitive architecture* assumptions. *Dual-Process Theory* offers a design grammar for distinguishing fast, heuristic pattern completion from slow, resource-intensive deliberation, which is directly actionable for interface pacing, explanation timing, and error-prevention (Pellert et al., 2024; Omarov, Narynov, et al., 2023). *Signal Detection Theory* translates model outputs into separable components of sensitivity and criterion placement, enabling explicit governance of false-alarm versus miss costs in high-stakes classification. *Drift Diffusion Model* intuitions map to confidence calibration, latency, and evidence accumulation, clarifying why seemingly persuasive outputs can still be epistemically underpowered when evidence quality is low. *Bounded Rationality* and *limited-capacity attentional control* specify why users cannot continuously audit AI, so systems must externalize uncertainty, constrain overconfident language, and design for cognitive offloading without creating cognitive atrophy.

Learning, Reinforcement Dynamics, and Habit Formation in Human-AI Feedback Systems

When AI systems personalize content, recommendations, or interventions, they enter the domain of *reinforcement learning* as a behavioral theory engine, even if they are not explicitly trained with reinforcement objectives. *Reward Prediction Error* provides a mechanistic account of how micro-rewards, intermittent reinforcement, and novelty signals shape persistence, craving-like loops, and attentional capture, which becomes operationally relevant for responsible engagement design (Chang et al., 2024; Saputra et al., 2024). The distinction between *model-based* control and *model-free* habit maps to autonomy-preserving design, since systems that maximize frictionless ease can shift users toward habitual responding, reducing reflective choice and undermining self-regulation. *Temporal discounting* and *delay sensitivity* clarify why short-term incentives can dominate long-term well-being, requiring explicit countermeasures such as commitment scaffolds, pacing constraints, and goal-congruent defaults. In psychologically sensitive contexts, feedback loops must be treated as endogenous causal systems, because predictions can change behavior, behavior changes data, and data then reifies the model's own worldview through performative reinforcement.

Memory, Attention, and Cognitive Control under Algorithmic Mediation

AI-mediated environments restructure cognition by reallocating attention, compressing memory demands, and reshaping executive control, which can be formalized through *working memory capacity* constraints and *cognitive load theory* principles. When interfaces externalize remembering through reminders, summaries, and auto-completions, they induce *cognitive offloading* that can improve functional performance while risking skill erosion in self-initiated retrieval, sustained attention, and strategic planning (Xu, 2023; Xia et al., 2023). Attentional dynamics are further modulated by algorithmic salience, where optimized novelty and emotional arousal act as exogenous drivers of *attentional capture*, increasing distraction, fatigue, and reduced depth processing. *Executive functions* such as inhibition, shifting, and updating become both targets and casualties of pervasive automation, so psychologically responsible systems must include workload-aware interaction cadence, interruption budgets, and explanation formats that respect limited attentional bandwidth. Concretely, model outputs should be throttled, uncertainty should be explicit, and user control should be structurally protected to prevent over-automation of metacognitive monitoring.

Language, Meaning, and Narrative Cognition in Generative and Conversational Systems

Generative and conversational AI operates directly on the substrate of language, which makes *psycholinguistics*, *pragmatics*, and *narrative psychology* central to understanding influence, misunderstanding, and identity-shaping effects. *Framing effects* and *cognitive schema activation* explain why semantically similar statements can produce different decisions, which requires governance of tone, certainty markers, and evaluative adjectives in high-stakes guidance (Liu et al., 2025; Cardon et al., 2023). *Discourse coherence* and *pragmatic implicature* clarify how users infer intent, authority, and hidden premises from conversational structure, so explanation design must minimize implicature-based overclaiming and avoid rhetorical closure that suppresses critical reflection. Narrative mechanisms matter because repeated interaction can shape *self-concept*, *autobiographical memory*, and *meaning-making*, especially when systems mirror user language in ways that validate maladaptive interpretations. Operationally, safe conversational systems must separate empathy signaling from epistemic authority, maintain epistemic humility, and implement boundary-aware prompting that supports user agency through options, not prescriptions.

Social Cognition, Mind Perception, and Embodied-Situated-Extended Frameworks for Human-AI Interaction

Human interaction with AI is governed by *social cognition* defaults, including *attribution theory*, *anthropomorphism*, and *mind perception*, which cause users to infer agency, intentionality, and moral status even when the system is mechanistically non-sentient. These inferences shape *trust calibration*, compliance, and blame allocation, producing predictable risks such as deference to confident outputs, diffusion of responsibility, and moral disengagement in delegated decision chains (Kong et al., 2024; Omarov, Zhumanov, et al., 2023). *Social identity theory* and norm psychology explain why AI-mediated ranking, visibility, and moderation can intensify in-group favoritism, status anxiety, and conformity pressures, especially in networked publics. At the same time, *embodied cognition* and *situated cognition* emphasize that meaning is action-context dependent, so the psychological impact of AI depends on affordances, constraints, and ecological cues, not merely content. *Extended mind* theory then frames AI tools as cognitive prostheses, implying design obligations to preserve competence, avoid dependency traps, and maintain contestability so users can reassert agency when automated guidance conflicts with lived context.

4. AI for Psychological Assessment, Measurement, and Inference

AI-driven psychological assessment must be conceptualized as a measurement system, not a prediction gadget, because psychological scores are interpretive claims about latent variables under error, context, and sampling constraints. *Classical Test Theory* foregrounds reliability as a lower bound on interpretability, while *Generalizability Theory* expands reliability into facet-specific variance components, such as rater, task, device, language, and setting (Ke et al., 2025; Dien, 2023). *Item Response Theory* and adaptive measurement logics translate into algorithmic item selection, difficulty calibration, and information optimization, but only when the target construct is specified with a defensible *nomological network*. *Structural Equation Modeling* clarifies that observed indicators are imperfect reflections of latent constructs, so feature embeddings must be treated as indicators with explicit residual structure, rather than as construct-equivalents. In this framing, *measurement invariance* is a governance requirement, ensuring that score meaning is stable across populations, cultures, neurotypes, and modalities, and that apparent differences are not artifacts of differential measurement functioning.

Moving from questionnaires to behavioral traces, passive sensing, and *digital phenotyping* expands temporal granularity and ecological proximity, yet it also intensifies construct ambiguity because traces are co-produced by platform affordances, incentive gradients, and normative self-presentation. Self-report retains privileged access to private experience, yet it is vulnerable to *social desirability*, recall compression, and introspective limits, which motivates *Ecological Momentary Assessment* as a compromise that captures within-person dynamics while reducing retrospection bias (Fan et al., 2023; Dolunay & Temel, 2024). Passive signals such as keystroke rhythms, mobility patterns, sleep proxies, and interaction latency can support state estimation, but only under a strict construct map that separates symptom correlates from socioeconomic constraints, disability access barriers, and device ecology. The core risk is proxy substitution, where convenience features become psychological labels, producing spurious inference and unjust allocation. A robust framework therefore requires context modeling, temporal anchoring of constructs, within-person baselining, and explicit rules against using convenience-correlated signals as if they were psychological essence.

Affective Computing, Emotion Inference, and the Measurement Problem of Expressivity

Emotion inference is an archetypal validity stress test because expression is not isomorphic with experience, and emotion categories are theory-dependent. *Basic Emotion* views imply discrete signatures, *Dimensional Models* emphasize valence-arousal gradients, *Appraisal Theories* prioritize situational meaning, and *Psychological Construction* treats emotion as an emergent categorization process, so an AI system must declare which ontology it assumes before claiming detection (Moin et al., 2025; Grassini, 2023). *Emotion Regulation* further decouples felt affect from displayed behavior through reappraisal, suppression, masking, and strategic presentation, making label assignment underdetermined without context. Cultural display rules, neurodiversity-related expressivity differences, and disability-mediated communication practices produce systematic measurement non-equivalence, which can be misread as risk, disengagement, or deceit (Yin et al., 2023; Seyfi et al., 2025). Affective pipelines must therefore implement uncertainty-aware outputs, abstention policies, and error-cost asymmetry governance, especially in high-stakes settings such as education, employment, and mental health triage. Without these constraints, affective computing becomes a high-confidence instrument for low-validity inference.

Trait, Motivation, Values, and Identity Inference under Algorithmic Profiling Pressures

Psychological profiling often collapses traits, states, and contexts into a single inferred profile, yet trait constructs are probabilistic dispositions expressed through person-situation transactions rather than invariant behavioral signatures. *Big Five* and *HEXACO* frameworks provide parsimonious trait taxonomies, but trait inference from digital traces is fragile because platform contexts can amplify situational strength and suppress behavioral variance (Alkahtani et al., 2024; Huo et al., 2025). Motivation inference requires more than preference prediction, since *Self-Determination Theory* distinguishes autonomy-supportive engagement from controlled compliance, and *Regulatory Focus* differentiates promotion versus prevention orientations that can look similar in clickstream data. Values and identity add further complexity, since *self-concept clarity* and narrative coherence are shaped by social feedback and institutional categorization, so algorithmic labels can become performative, altering the very construct they purport to measure (Shahid et al., 2024; Kjell et al., 2024). Responsible inference therefore demands strict purpose limitation, explicit construct boundaries, and safeguards against reification, including contestability, time-bounded profiles, and constraints against using inferred traits for exclusionary decisions without transparent justification and human review.

In psychologically consequential systems, the central design task is not merely maximizing accuracy, it is governing validity threats and decision thresholds under asymmetric harms. Proxy confounding can arise when models learn socioeconomic correlates, language registers, or institutional frictions that masquerade as psychological states, while *dataset shift* and *concept drift* can silently degrade performance as cultures, platforms, and norms evolve (Li et al., 2024; Teng et al., 2024). Feedback loops create performativity, where predictions reshape behavior, behavior reshapes data, and the loop stabilizes a self-confirming narrative. Interpretability can also be illusory, because plausible explanations can satisfy human coherence heuristics while misrepresenting actual feature influence, amplifying overtrust and compliance (Carolus et al., 2023; Rane et al., 2024). Threshold selection must therefore be framed through *signal detection* trade-offs and explicit harm modeling, specifying acceptable false-alarm and miss rates by context, and embedding abstention, escalation, and recourse pathways. A psychologically safe architecture couples conservative decision policies with monitoring for drift, subgroup non-equivalence, and downstream adverse effects such as stigma amplification, avoidance reinforcement, and diminished help-seeking.

5. AI-Mediated Intervention, Psychotherapy Translation, and Behavior Change Systems

AI-mediated intervention should be treated as a translation problem from therapeutic theory into interactional affordances, where psychological efficacy depends on mechanism fidelity, ethical boundary conditions, and context-aware implementation. *Cognitive Behavioral Therapy* offers modularizable components such as cognitive restructuring, behavioral activation, graded exposure, and skills rehearsal, yet each component requires careful operationalization to avoid superficial advice-giving that bypasses formulation and context (Ladak et al., 2024; Cao et al., 2023). *Acceptance and Commitment Therapy* frames change through values clarification, experiential acceptance, and cognitive defusion, which implies an autonomy-supportive conversational stance, minimizing prescriptive language and supporting self-authored commitments. *Dialectical Behavior Therapy* specifies crisis-relevant skills such as distress tolerance and emotion regulation, which demands hard safety constraints, escalation pathways, and non-coercive crisis communication protocols (Landers & Behrend, 2023; Xu et al., 2023). *Psychodynamic* and relational paradigms emphasize meaning, defenses, and transference-like dynamics, implying that simulated intimacy can induce attachment-like reliance and interpretive overreach if the system performs empathy without clinical accountability. *Humanistic* approaches prioritize authenticity, unconditional positive regard, and empathic attunement, which require epistemic humility and strict separation between validation and diagnostic authority.

Intervention credibility is maximized when design aligns with mechanisms of change, not merely symptom suppression, because psychological outcomes are mediated by cognitive, affective, behavioral, and interpersonal processes. *Common Factors* theory foregrounds expectancy, goal consensus, perceived support, and alliance-like engagement, which can be partially instantiated through consistent tone, transparent boundaries, and user-centered goal articulation (Srinivasan & Rajavel, 2025; Guingrich & Graziano, 2024). Process-based approaches emphasize transdiagnostic mechanisms such as avoidance, rumination, cognitive fusion, threat monitoring, and emotion dysregulation, enabling interventions to target processes that generalize across diagnostic categories. *Self-efficacy* and mastery experiences become both mechanisms and outcomes, implying that over-automation can be counterproductive by reducing opportunities for effortful skill acquisition and reflective learning (Timmons et al., 2023; Uwaifo & Uwaifo, 2023). *Inhibitory learning* principles in exposure-based change require graded uncertainty, tolerable arousal, and reconsolidation-compatible pacing, which cannot be replaced by reassurance or distraction. A rigorous design stance therefore requires mechanism mapping, proximal process indicators, and failure-mode anticipation, including iatrogenic reinforcement of avoidance, dependence formation, and stigma internalization.

Behavior Change Architectures, Choice Engineering, and Autonomy-Supportive Persuasion Boundaries

AI systems that recommend, nudge, or coach inevitably implement a theory of behavior change, whether explicitly or covertly, so ethical design requires making that theory legible and autonomy-compatible. *Theory of Planned Behavior* and expectancy-value logics clarify that intention formation depends on attitudes, perceived norms, and perceived control, implying that interventions must increase capability and agency rather than merely intensify persuasion. *Protection Motivation* and risk perception frameworks suggest that fear-based messaging can backfire through defensiveness and avoidance unless paired with efficacy and feasible action pathways (Chen et al., 2024; Shen & Cui, 2024). Stage-based change models emphasize readiness heterogeneity, warning against uniform prompts that provoke reactance among low-readiness users. Capability-opportunity-motivation framing and behavioral systems thinking support constraint mapping, identifying structural barriers, time scarcity, and social risk that make individual-level nudges ineffective or unjust (Valový & Buchalcevova, 2023; Bazarkina, 2023). *Nudge theory* and choice architecture must therefore be bounded by autonomy-supportive constraints, including informed consent, opt-out friction symmetry, and prohibitions on dark-pattern persuasion that exploits vulnerability, scarcity cognition, or social comparison anxiety.

Engagement, Adherence, and the Psychodynamics of Sustained Use in Digital Interventions

Sustained engagement is not a neutral technical metric, it is a psychological phenomenon shaped by motivation quality, reinforcement contingencies, identity alignment, and perceived respect for agency. *Self-Determination Theory* predicts that competence support, autonomy support, and relatedness cues can foster durable engagement, while controlling prompts and surveillance cues can produce disengagement, oppositionality, or performative compliance (Ghosh, 2024; Jin & Zhang, 2025). *Flow theory* reframes personalization as a calibration problem between challenge and skill, implying that adaptive difficulty, scaffolded mastery, and feedback specificity can sustain engagement without resorting to compulsive reinforcement. At the same time, variable reinforcement schedules and novelty optimization can induce habit loops that mimic addictive dynamics, requiring explicit ethical boundaries and engagement caps in psychologically sensitive tools (Lau et al., 2025; Bergdahl et al., 2023). Adherence is also shaped by cognitive load and attentional bandwidth, so micro-interventions must be parsimonious, context-aware, and aligned with real-world routines rather than idealized schedules. A psychologically responsible design therefore treats engagement as value-congruent participation, not as a manipulable retention target, and builds disengagement pathways that preserve dignity and encourage appropriate human help-seeking when needed.

Safety-by-Design, Crisis Boundaries, and Iatrogenic Risk Management in AI-Mediated Care

Safety in AI-mediated psychological intervention must be engineered as an integrated risk system, because harms can emerge from misclassification, misguidance, overdependence, or delayed escalation rather than from overtly harmful content alone. Crisis contexts demand conservative decision policies, including abstention when uncertainty is high, rapid escalation to human support, and carefully constrained language that avoids false reassurance or coercive directives (Kim et al., 2025; Crawford et al., 2024). Iatrogenic risks include reinforcement of avoidance through premature soothing, escalation of rumination through excessive problem elaboration, identity threat through labeling, and reduced help-seeking through perceived sufficiency of automated support. Dependency dynamics are amplified by anthropomorphic design, consistent availability, and intimacy cues, which can shift relational needs toward the system and erode real-world support-seeking (Hui et al., 2024; Bonnefon et al., 2024). Effective safety governance therefore requires boundary transparency, explicit competence limits, and interactional guardrails that reduce moral hazard, including non-diagnostic framing, harm-aware prompting, and structured pathways for professional referral. Post-deployment safety also requires continuous monitoring for drift in user populations, emergent misuse, and unintended behavioral externalities, because psychological systems are adaptive, context-dependent, and socially contagious.

6. Societal, Organizational, and Policy Dimensions of AI in Psychology

Equity in psychologically consequential AI is a multilayer construct spanning distributive, procedural, and recognitional justice, where harms often arise from construct misalignment rather than overt intent. *Intersectionality* and *minority stress* frameworks clarify that risk is cumulative and non-linear, because exposure to stigma, surveillance, and constrained opportunity structures can amplify threat vigilance, affective dysregulation, and disengagement (Salah et al., 2023; Alkhouri, 2024). *Cultural models of self* and culturally patterned emotion norms further destabilize universalist inference, since the same behavioral trace can encode divergent meanings across collectivist-individualist ecologies, high-context communication regimes, and disability-mediated interaction styles. Fairness governance must therefore move beyond parity metrics toward construct-level comparability, including measurement invariance, language justice, and context-sensitive interpretability (Olawade et al., 2024; Kediya et al., 2023). Psychologically informed auditing also requires attention to *stereotype threat*, *identity contingency*, and differential error costs, because false positives can trigger institutional exclusion, while false negatives can delay care, both producing downstream psychosocial scarring.

Privacy, Autonomy, and Psychological Integrity in Datafied Life Worlds

Privacy in AI-mediated psychological systems is not merely a compliance checkbox, it is a prerequisite for autonomy, identity exploration, and affective safety, because selfhood is partly constructed through controlled disclosure and context-specific self-presentation. *Contextual integrity* and informational self-determination map privacy to legitimate information flows, purpose limitation, and consent that is cognitively comprehensible rather than nominal (Obidovna, 2024; De Freitas et al., 2023). When systems infer states such as stress, vulnerability, or risk from passive traces, the threat shifts from data exposure to inference exposure, creating a distinctive mental privacy problem that can generate chilling effects, anticipatory anxiety, and strategic self-censorship. Autonomy degradation also occurs through covert personalization, where choice architectures exploit attentional vulnerabilities, reward sensitivity, and social comparison pressures, producing compliance without reflective endorsement. Psychological integrity requires governance that protects refusal rights, minimizes data, separates supportive functions from evaluative surveillance, and constrains downstream secondary use, because secondary use converts supportive environments into disciplinary infrastructures that erode trust, help-seeking, and social belonging.

Organizational Adoption, Workforce Psychodynamics, and Implementation Failure Modes

Organizational deployment of psychologically consequential AI is best modeled as a socio-technical transition that reshapes roles, authority gradients, and epistemic norms, rather than as a tool installation. *Job Demands-Resources* logic predicts that AI can either reduce cognitive load through decision support or intensify strain through pace acceleration, monitoring externalities, and documentation burdens that deplete recovery resources (Eid et al., 2023; Shiffrin & Mitchell, 2023). *Psychological safety* becomes a control variable for harm detection, since low safety climates suppress dissent, normalize automation deference, and delay incident reporting until damage becomes systemic.

Professional identity dynamics also matter, because clinicians, educators, and managers may experience de-skilling anxiety, moral distress, or status threat when algorithmic recommendations are framed as superior, producing resistance, covert workarounds, or overreliance (Liu & Chang, 2024; Martin & Zimmermann, 2024). Implementation must therefore include workload-sensitive workflow design, role-clarity, contestability channels, and competence-building, while avoiding performative compliance regimes. Without these safeguards, adoption produces brittle systems, silent errors, and institutionalized blame shifting that undermines trust and long-term effectiveness.

Accountability, Governance, and the Ethics of Delegation in High-Stakes Psychotechnical Systems

Accountability in AI-psychology interfaces requires explicit allocation of decision rights, evidentiary standards, and liability pathways across designers, deployers, and institutions, because diffusion of responsibility is a predictable outcome of complex automation chains. *Procedural justice* principles translate into contestability, intelligibility, and due-process guarantees, where affected individuals can understand the basis of a decision, challenge it, and obtain meaningful recourse (Ke et al., 2025; Obidovna, 2024). Governance must operationalize lifecycle discipline, including pre-deployment risk registers, abuse-case modeling, uncertainty thresholds, post-deployment monitoring for drift, and incident response protocols that treat psychological harm as a reportable safety event rather than an anecdotal complaint. Auditability should be designed into the system through documentation of data provenance, construct definitions, performance boundaries, and change logs, enabling institutional learning rather than retrospective scapegoating (Kim et al., 2025; Alkhouri, 2024). Ethical delegation also requires constraints on automation scope, because substituting human judgment with algorithmic authority in contexts involving dignity, stigma risk, or coercive power can produce legitimacy crises, reduced trust, and adverse behavioral adaptation even when nominal accuracy remains high.

Education, Professional Formation, and Future-Oriented Capacity Building

A durable global response to AI in psychology depends on competency architectures that integrate measurement literacy, human factors, ethics-by-design, and governance fluency into professional formation across clinical, educational, organizational, and policy domains. *Metacognitive calibration* must be treated as a trainable skill, because appropriate reliance requires understanding uncertainty, base rates, and error costs, not merely learning tool features (Ladak et al., 2024; Uludag, 2025). Interdisciplinary collaboration is also a structural requirement, with boundary objects such as construct definitions, harm taxonomies, and evaluation protocols enabling coordination across psychology, data science, design, law, and public administration. Workforce development should prioritize translational competence, including the ability to convert theories such as *self-determination*, *cognitive load*, and *social influence* into system requirements, interface constraints, and procurement criteria (Gigerenzer, 2024; Bonnefon et al., 2024). Future-oriented capacity building additionally requires institutional infrastructures for continuous auditing, red-team style failure anticipation, and public-facing accountability, ensuring that psychological integrity, equity, and autonomy remain first-class objectives as models evolve, scale, and diffuse across cultures, languages, and governance regimes.

7. Conclusion

This article has treated AI in psychological domains as a construct-governed socio-technical phenomenon, where epistemic legitimacy depends on theory-consistent measurement, mechanism-aware intervention logic, and institutionally enforceable safeguards. The throughline is that psychological inference is never a mere computational convenience, because constructs such as emotion, motivation, trust, and identity are context-saturated, norm-mediated, and dynamically co-produced through interaction. When AI systems are deployed as assessors, recommenders, or quasi-therapeutic agents, they become part of the causal ecology that shapes cognition, affect, and social behavior, which renders naive predictionism insufficient. A coherent integrative stance therefore couples *construct validity* with *ecological validity*, aligns predictive surfaces with mechanism plausibility, and treats uncertainty as a communicative obligation rather than a technical residual. The article has also foregrounded that psychological safety is a design target, requiring cognitive-ergonomic interfaces, autonomy-supportive interaction patterns, and harm-aware thresholds that respect error-cost asymmetry. Across sectors, the decisive competence is translational capacity, the ability to map psychological theories into system requirements, governance controls, and contestable decision architectures.

A submission-ready set of principles emerges when psychological constructs are treated as hypotheses embedded in a *nomological network*, rather than as labels discoverable from traces. The first principle is *construct discipline*, specifying the target construct, admissible indicators, and boundary conditions, while mandating invariance checks across cultures, languages, neurotypes, and socioeconomic ecologies. The second principle is *calibration governance*, requiring explicit uncertainty representation, abstention when evidence is weak, and interface designs that reduce *automation bias* and preserve user metacognition. The third principle is *autonomy protection*, grounded in *self-determination theory* and procedural justice, ensuring that personalization does not mutate into coercive choice architecture, surveillance, or dependency-inducing pseudo-relationship design. The fourth principle is *mechanism fidelity*, aligning AI-mediated intervention with process-based change models, common factors, and iatrogenic risk controls, so systems support mastery, self-efficacy, and reflective agency rather than mere compliance. The fifth principle is *institutional accountability*, embedding auditability, contestability, and lifecycle monitoring for drift, feedback loops, and adverse psychosocial externalities, so responsibility remains allocatable and harms are governable.

Global readiness requires that psychological expertise is operationalized into governance, procurement, and workforce development, because AI diffusion is outpacing institutional capacity to evaluate psychological claims and control downstream harms. For academic ecosystems, the agenda is to stabilize shared conceptual standards for construct mapping, explanation legitimacy, and intervention boundaries, enabling cumulative theory-to-design translation without overclaiming. For industry and digital learning technologists, the actionable mandate is to engineer psychologically literate product architectures, including friction symmetry, uncertainty-first outputs, crisis-aware escalation, and context-respecting personalization that treats user dignity as a non-negotiable constraint.

For policymakers and public administrators, the critical move is to treat *psychological integrity* as a protected interest, enforceable through audit obligations, contestability rights, surveillance limits, and proportionality in high-stakes decision automation. For workforce development professionals, the core task is competency formation, building measurement literacy, human factors fluency, causal reasoning, and ethics-by-design capacity across multidisciplinary teams. If these cross-sector capacities mature in tandem, AI can function as a bounded augmentation of human psychological practice, while reducing predictable harms such as stigma amplification, autonomy erosion, and institutionally normalized mis-measurement.

Declaration of Interest:
No potential conflict of interest was reported by the authors.

Funding Information:
This research did not receive any specific funding from any public, commercial, or non-profit agency.

Disclosure Statement:
No material or relevant stake relating to this research was disclosed by the author(s).

Competing Interest:
No potential conflict of interest was reported by the author(s).

Data Availability Statement:
Data sharing is not applicable to this research article as no new data were created or analysed in this study.

References

- Alkahtani, H., Aldhyani, T. H., & Alqarni, A. A. (2024). Artificial intelligence models to predict disability for mental health disorders. *Journal of Disability Research, 3*(3), 20240022. https://doi.org/10.57197/JDR-2024-0022
- Alkhouri, K. I. (2024). The role of artificial intelligence in the study of the psychology of religion. *Religions, 15*(3), 290. https://doi.org/10.3390/rel15030290
- Al-Mseidin, K. I., Al-Zu'bi, M. A., Al-Mawadieh, R. S., & Alelaimat, A. M. (2023, December). The integration of AI applications in psychological and educational counselling for university students: Opportunities and challenges. In *2023 24th International Arab Conference on Information Technology (ACIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ACIT58888.2023.10453762
- Bazarkina, D. (2023). Current and future threats of the malicious use of artificial intelligence by terrorists: Psychological aspects. In *The Palgrave handbook of malicious use of AI and psychological security* (pp. 251–272). Springer International Publishing. https://doi.org/10.1007/978-3-031-22552-9_10
- Bergdahl, J., Latikka, R., Celuch, M., Savolainen, I., Mantere, E. S., Savela, N., & Oksanen, A. (2023). Self-determination and attitudes toward artificial intelligence: Cross-national and longitudinal perspectives. *Telematics and Informatics, 82*, 102013. https://doi.org/10.1016/j.tele.2023.102013
- Bonnefon, J. F., Rahwan, I., & Shariff, A. (2024). The moral psychology of artificial intelligence. *Annual Review of Psychology, 75*(1), 653–675. https://doi.org/10.1146/annurev-psych-030123-113559
- Cao, L., Chen, C., Dong, X., Wang, M., & Qin, X. (2023). The dark side of AI identity: Investigating when and why AI identity entitles unethical behavior. *Computers in Human Behavior, 143*, 107669. https://doi.org/10.1016/j.chb.2023.107669
- Cardon, P. W., Ma, H., & Fleischmann, C. (2023). Recorded business meetings and AI algorithmic tools: Negotiating privacy concerns, psychological safety, and control. *International Journal of Business Communication, 60*(4), 1095–1122. https://doi.org/10.1177/23294884211037009
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS–Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014
- Chang, P. C., Zhang, W., Cai, Q., & Guo, H. (2024). Does AI-driven technostress promote or hinder employees' artificial intelligence adoption intention? A moderated mediation model of affective reactions and technical self-efficacy. *Psychology Research and Behavior Management, 17*, 413–427. https://doi.org/10.2147/PRBM.S441444

- Chen, J., Yuan, D., Dong, R., Cai, J., Ai, Z., & Zhou, S. (2024). Artificial intelligence significantly facilitates development in the mental health of college students: A bibliometric analysis. *Frontiers in Psychology, 15*, 1375294. https://doi.org/10.3389/fpsyg.2024.1375294
- Crawford, J., Allen, K. A., Pani, B., & Cowling, M. (2024). When artificial intelligence substitutes humans in higher education: The cost of loneliness, student success, and retention. *Studies in Higher Education, 49*(5), 883–897. https://doi.org/10.1080/03075079.2024.2326956
- De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour, 7*(11), 1845–1854. https://doi.org/10.1038/s41562-023-01734-2
- Dien, J. (2023). Generative artificial intelligence as a plagiarism problem. *Biological Psychology, 181*, 108621. https://doi.org/10.1016/j.biopsycho.2023.108621
- Dolunay, A., & Temel, A. C. (2024). The relationship between personal and professional goals and emotional state in academia: A study on unethical use of artificial intelligence. *Frontiers in Psychology, 15*, 1363174. https://doi.org/10.3389/fpsyg.2024.1363174
- Eid, M. M., Yundong, W., Mensah, G. B., & Pudasaini, P. (2023). Treating psychological depression utilising artificial intelligence: AI for precision medicine—Focus on procedures. *Mesopotamian Journal of Artificial Intelligence in Healthcare*, 76–81. https://doi.org/10.58496/MJAIH/2023/015
- Fan, X., Zhao, S., Zhang, X., & Meng, L. (2023). The impact of improving employee psychological empowerment and job performance based on deep learning and artificial intelligence. *Journal of Organizational and End User Computing, 35*(3), 1–14. https://doi.org/10.4018/JOEUC.321639
- Ghosh, S. (2024). Artificial intelligence in future psychological revolution. In *Mind and machines: The psychology of artificial intelligence* (p. 113).
- Gigerenzer, G. (2024). Psychological AI: Designing algorithms informed by human psychology. *Perspectives on Psychological Science, 19*(5), 839–848. https://doi.org/10.1177/17456916231180597
- Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology, 14*, 1191628. https://doi.org/10.3389/fpsyg.2023.1191628
- Guingrich, R. E., & Graziano, M. S. (2024). Ascribing consciousness to artificial intelligence: Human-AI interaction and its carry-over effects on human-human interaction. *Frontiers in Psychology, 15*, 1322781. https://doi.org/10.3389/fpsyg.2024.1322781
- Gupta, A., & Guglani, A. (2023). Scenario analysis of malicious use of artificial intelligence and challenges to psychological security in India. In *The Palgrave handbook of malicious use of AI and psychological security* (pp. 397–418). Springer International Publishing. https://doi.org/10.1007/978-3-031-22552-9_15
- Harding, J., D'Alessandro, W., Laskowski, N. G., & Long, R. (2024). AI language models cannot replace human research participants. *AI & Society, 39*(5), 2603–2605. https://doi.org/10.1007/s00146-023-01725-x
- Hui, Z., Khan, A. N., Chenglong, Z., & Khan, N. A. (2024). When service quality is enhanced by human–artificial intelligence interaction: An examination of anthropomorphism and responsiveness from the perspectives of employees and customers. *International Journal of Human-Computer Interaction, 40*(22), 7546–7561. https://doi.org/10.1080/10447318.2023.2266254
- Huo, W., Li, Q., Liang, B., Wang, Y., & Li, X. (2025). When healthcare professionals use AI: Exploring work well-being through psychological needs satisfaction and job complexity. *Behavioral Sciences, 15*(1), 88. https://doi.org/10.3390/bs15010088
- Jin, F., & Zhang, X. (2025). Artificial intelligence or human: When and why consumers prefer AI recommendations. *Information Technology & People, 38*(1), 279–303. https://doi.org/10.1108/ITP-01-2023-0022
- Ke, L., Tong, S., Cheng, P., & Peng, K. (2025). Exploring the frontiers of LLMs in psychological applications: A comprehensive review. *Artificial Intelligence Review, 58*(10), 305. https://doi.org/10.1007/s10462-025-11297-5
- Kediya, S. O., Dhote, S., Singh, D. K., Bidve, V. S., Pathan, S., & Suchak, A. (2023). Are AI and chat bot services affecting the psychology of users in banking services and financial sector. *Journal for ReAttach Therapy and Developmental Diversities, 6*(2), 191–197.
- Kim, J. J., Soh, J., Kadkol, S., Solomon, I., Yeh, H., Srivatsa, A. V., & Ajilore, O. (2025). AI anxiety: A comprehensive analysis of psychological factors and interventions. *AI and Ethics*, 1–17. https://doi.org/10.1007/s43681-025-00686-9
- Kjell, O. N., Kjell, K., & Schwartz, H. A. (2024). Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment. *Psychiatry Research, 333*, 115667. https://doi.org/10.1016/j.psychres.2023.115667
- Kong, H., Yin, Z., Chon, K., Yuan, Y., & Yu, J. (2024). How does artificial intelligence (AI) enhance hospitality employee innovation? The roles of exploration, AI trust, and proactive personality. *Journal of Hospitality Marketing & Management, 33*(3), 261–287. https://doi.org/10.1080/19368623.2023.2258116
- Ladak, A., Loughnan, S., & Wilks, M. (2024). The moral psychology of artificial intelligence. *Current Directions in Psychological Science, 33*(1), 27–34. https://doi.org/10.1177/09637214231205866
- Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist, 78*(1), 36–49. https://doi.org/10.1037/amp0000972
- Lau, Y., Ang, W. H. D., Ang, W. W., Pang, P. C. I., Wong, S. H., & Chan, K. S. (2025). Artificial intelligence-based psychotherapeutic intervention on psychological outcomes: A meta-analysis and meta-regression. *Depression and Anxiety, 2025*(1), 8930012. https://doi.org/10.1155/da/8930012

- Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human-AI trust. *Frontiers in Psychology, 15*, 1382693. https://doi.org/10.3389/fpsyg.2024.1382693
- Liu, H., Fan, J., & Xia, M. (2025). Exploring individuals' emotional and autonomous learning profiles in AI-enhanced data-driven language learning: An expanded SOR perspective. *Learning and Individual Differences, 122*, 102753. https://doi.org/10.1016/j.lindif.2025.102753
- Liu, Y., & Chang, P. (2024). Exploring EFL teachers' emotional experiences and adaptive expertise in the context of AI advancements: A positive psychology perspective. *System, 126*, 103463. https://doi.org/10.1016/j.system.2024.103463
- Martin, K. D., & Zimmermann, J. (2024). Artificial intelligence and its implications for data privacy. *Current Opinion in Psychology, 58*, 101829. https://doi.org/10.1016/j.copsyc.2024.101829
- Marx, E., Leonhardt, T., & Bergner, N. (2023). Secondary school students' mental models and attitudes regarding artificial intelligence: A scoping review. *Computers and Education: Artificial Intelligence, 5*, 100169. https://doi.org/10.1016/j.caeai.2023.100169
- Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports, 14*(1), 4692. https://doi.org/10.1038/s41598-024-53755-0
- Moin, M. F., Behl, A., Zhang, J. Z., & Shankar, A. (2025). AI in the organizational nexus: Building trust, cementing commitment, and evolving psychological contracts. *Information Systems Frontiers, 27*(4), 1413–1424. https://doi.org/10.1007/s10796-024-10561-3
- Obidovna, D. Z. (2024). The pedagogical-psychological aspects of artificial intelligence technologies in integrative education. *International Journal of Literature and Languages, 4*(3), 13–19. https://doi.org/10.37547/ijll/Volume04Issue03-03
- Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with artificial intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health, 3*, 100099. https://doi.org/10.1016/j.glmedi.2024.100099
- Omarov, B., Narynov, S., & Zhumanov, Z. (2023). Artificial intelligence-enabled chatbots in mental health: A systematic review. *Computers, Materials & Continua, 74*(3). https://doi.org/10.32604/cmc.2023.034655
- Omarov, B., Zhumanov, Z., Gumar, A., & Kuntunova, L. (2023). Artificial intelligence-enabled mobile chatbot psychologist using AIML and cognitive behavioral therapy. *International Journal of Advanced Computer Science and Applications, 14*(6). https://doi.org/10.14569/IJACSA.2023.0140616
- Paesano, A. (2023). Artificial intelligence and creative activities inside organizational behavior. *International Journal of Organizational Analysis, 31*(5), 1694–1723. https://doi.org/10.1108/IJOA-09-2020-2421
- Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., & Strohmaier, M. (2024). AI psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science, 19*(5), 808–826. https://doi.org/10.1177/17456916231214460
- Piotrowski, C. (2025). Artificial intelligence (AI) and its potential in personality assessment and projective testing. *SIS Journal of Projective Psychology & Mental Health, 32*(1).
- Presbitero, A., & Teng-Calleja, M. (2023). Job attitudes and career behaviors relating to employees' perceived incorporation of artificial intelligence in the workplace: A career self-management perspective. *Personnel Review, 52*(4), 1169–1187. https://doi.org/10.1108/PR-02-2021-0103
- Priya, B., & Sharma, V. (2023). Exploring users' adoption intentions of intelligent virtual assistants in financial services: An anthropomorphic and socio-psychological perspective. *Computers in Human Behavior, 148*, 107912. https://doi.org/10.1016/j.chb.2023.107912
- Rane, N., Choudhary, S., & Rane, J. (2024). Artificial intelligence for enhancing resilience. *Journal of Applied Artificial Intelligence, 5*(2), 1–33. https://doi.org/10.48185/jaai.v5i2.1053
- Salah, M., Al Halbusi, H., & Abdelfattah, F. (2023). May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Computers in Human Behavior: Artificial Humans, 1*(2), 100006. https://doi.org/10.1016/j.chbah.2023.100006
- Samoilenko, S. A., & Suvorova, I. (2023). Artificial intelligence and deepfakes in strategic deception campaigns: The US and Russian experiences. In *The Palgrave handbook of malicious use of AI and psychological security* (pp. 507–529). Springer International Publishing. https://doi.org/10.1007/978-3-031-22552-9_19
- Saputra, F. E., Buhalis, D., Augustyn, M. M., & Marangos, S. (2024). Anthropomorphism-based artificial intelligence (AI) robots typology in hospitality and tourism. *Journal of Hospitality and Tourism Technology, 15*(5), 790–807. https://doi.org/10.1108/JHTT-03-2024-0171
- Schwesig, R., Brich, I., Buder, J., Huff, M., & Said, N. (2023). Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people's willingness to use AI. *Journal of Risk Research, 26*(10), 1053–1084. https://doi.org/10.1080/13669877.2023.2249927
- Seyfi, S., Kim, M. J., Lee, C., Jo, Y., & Zaman, M. (2025). Exploring functional and psychological barriers to generative AI adoption for travel: A cross-cultural study. *Journal of Travel Research.* https://doi.org/10.1177/00472875251332955
- Shahid, M. K., Zia, T., Bangfan, L., Iqbal, Z., & Ahmad, F. (2024). Exploring the relationship of psychological factors and adoption readiness in determining university teachers' attitude on AI-based assessment systems. *The International Journal of Management Education, 22*(2), 100967. https://doi.org/10.1016/j.ijme.2024.100967
- Shen, Y., & Cui, W. (2024). Perceived support and AI literacy: The mediating role of psychological needs satisfaction. *Frontiers in Psychology, 15*, 1415248. https://doi.org/10.3389/fpsyg.2024.1415248

- Shiffrin, R., & Mitchell, M. (2023). Probing the psychology of AI models. *Proceedings of the National Academy of Sciences, 120*(10), e2300963120. https://doi.org/10.1073/pnas.2300963120
- Srinivasan, S., & Rajavel, N. (2025). Artificial intelligence approach to psychological wellbeing among the ageing population. In *AI technologies and advancements for psychological well-being and healthcare* (pp. 187–218). IGI Global. https://doi.org/10.4018/979-8-3693-9158-7.ch008
- Tassiello, V., Amatulli, C., Tillotson, J. S., & Laker, B. (2025). aiWOM: Artificial intelligence word-of-mouth—Conceptualizing consumer-to-AI communication. *International Journal of Human-Computer Interaction, 41*(7), 4248–4260.
- Teng, R., Zhou, S., Zheng, W., & Ma, C. (2024). Artificial intelligence (AI) awareness and work withdrawal: Evaluating chained mediation through negative work-related rumination and emotional exhaustion. *International Journal of Contemporary Hospitality Management, 36*(7), 2311–2326. https://doi.org/10.1108/IJCHM-02-2023-0240
- Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., … Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science, 18*(5), 1062–1096. https://doi.org/10.1177/17456916221134490
- Uludag, K. (2025). The use of AI-supported chatbot in psychology. In *Chatbots and mental healthcare in psychology and psychiatry* (pp. 1–20). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3112-5.ch001
- Uwaifo, F., & Uwaifo, A. O. (2023). Bridging the gap in alcohol use disorder treatment: Integrating psychological, physical, and artificial intelligence interventions. *International Journal of Applied Research in Social Sciences, 5*(4), 1–9.
- Valový, M., & Buchalcevova, A. (2023, October). The psychological effects of AI-assisted programming on students and professionals. In *2023 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 385–390). IEEE. https://doi.org/10.1109/ICSME58846.2023.00050
- Watermann, L., Kubowitsch, S., & Lermer, E. (2025). AI and work design: A positive psychology approach to employee well-being. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie*, 1–10. https://doi.org/10.1007/s11612-025-00806-3
- Xia, Q., Chiu, T. K., & Chai, C. S. (2023). The moderating effects of gender and need satisfaction on self-regulated learning through artificial intelligence (AI). *Education and Information Technologies, 28*(7), 8691–8713. https://doi.org/10.1007/s10639-022-11547-x
- Xu, G., Xue, M., & Zhao, J. (2023). The association between artificial intelligence awareness and employee depression: The mediating role of emotional exhaustion and the moderating role of perceived organizational support. *International Journal of Environmental Research and Public Health, 20*(6), 5147. https://doi.org/10.3390/ijerph20065147
- Xu, Y. (2023). What can AI learn from psychology and when can AI neglect it? An ontogeny-free study on the cognition of numerosity. *Fudan Journal of the Humanities and Social Sciences, 16*(4), 495–513. https://doi.org/10.1007/s40647-023-00381-1
- Yin, D., Li, M., & Qiu, H. (2023). Do customers exhibit engagement behaviors in AI environments? The role of psychological benefits and technology readiness. *Tourism Management, 97*, 104745. https://doi.org/10.1016/j.tourman.2023.104745