

Review

Psychological and Behavioral Modeling: A Structured Review of Methods, Datasets, Applications, and Open Challenges

Xingxuan Chen ^{1,†}, Xinmeng Li ^{2,†}, Lujia Yin ¹, Haixiang Zhu ², Zixin Xiao ² and Yong Peng ^{1,*} 

¹ College of Systems Engineering, National University of Defense Technology, Changsha 410073, China; chenxingxuan20@nudt.edu.cn (X.C.)

² Hunan Institute of Advanced Technology, Changsha 410205, China; lixinmeng15@nudt.edu.cn (X.L.)

* Correspondence: yongpeng@nudt.edu.cn; Tel.: +86-10-13973135400

† These authors contributed equally to this work.

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- We contribute a systems-level framework that organizes psychological and behavioral modeling across nested levels of analysis (individual, dyad and small group, group, with the societal level as an extension) and five modeling classes defined by unit of analysis, modeling target, and direction of inference; we place system dynamics, in the stock–flow–feedback sense used in this Special Issue, explicitly within that framework and distinguish it from dynamical-systems modeling in general.
- We contribute to systems methodology by comparing model families on assumptions, data requirements, interpretability, and failure modes rather than on non-comparable accuracy figures; by operationalizing the computational readiness of psychological theories on two rubric-scored axes; and by treating the micro–macro aggregation operator (mean-field, network, agent-based, or stock–flow) as an explicit and testable modeling choice.

What are the main findings and/or the implications of the main findings?

- Computational readiness, defined as the extent to which a theory can be implemented in a model without adding assumptions the theory does not state, and not algorithmic capability, determines where theory-informed learning is currently feasible; along the ladder from individual to collective units the dominant use of models shifts from estimation validated against the same units' data to simulation validated by macro-pattern reproduction, and validation practice has not followed that shift.
- The binding constraint on cross-scale modeling is multilevel data rather than model capacity, and the evidence base is heterogeneous in quantity and methodological quality, so claims of readiness must be indexed to a stated evidence tier rather than to benchmark scores. Each of the four claims is given explicit boundary conditions and the observations that would falsify it.



Academic Editor: Gianfranco Minati

Received: 9 August 2026

Revised: 21 September 2026

Accepted: 10 October 2026

Published: 19 October 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.

Abstract

Psychological and behavioral modeling turns cognition, emotion, and motivation into computable quantities that can be related to observable behavior. The field has moved from theory-driven rules to statistical machine learning, deep learning, large language models (LLMs), and foundation models of cognition, but methods are fragmented, results are seldom comparable, and the link between opaque predictors and psychological theory is unsettled. This structured narrative review, reported with a PRISMA-style selection flow

and an explicit evidence-tier coding, frames the field as a multilevel system. It organizes research into five modeling classes (psychological states, behavioral capabilities, individual–environment interaction, individual dynamics, and group psychology), defined by unit of analysis, target, and direction of inference, and it positions system dynamics, in the stock–flow–feedback sense, as one model family within that scheme rather than as a synonym for dynamical-systems modeling. We score the computational readiness of the theories that supply model ontologies on two rubric-based axes, formal specification and measurement maturity, and compare model families by assumptions, data needs, interpretability, and failure modes. Four claims are defended, each with stated boundary conditions and counterexamples: theory readiness, not algorithmic capability, determines where theory-informed learning is feasible; the dominant use of models shifts from individual-level estimation to collective-level simulation without a parallel shift in validation; multilevel data, not model capacity, constrain cross-scale modeling; and readiness claims must be indexed to a graded evidence base. We catalogue datasets, including small-group, dyadic, and longitudinal-network resources, review applications in mental health, education, decision support, and emergency management with their evidentiary status made explicit, treat military use as an ethically bounded case, and turn five recurrent bottlenecks into research-practice recommendations paired, where possible, with testable predictions.

Keywords: psychological and behavioral modeling; systems science; system dynamics; multilevel modeling; computational psychology; theory-informed machine learning; model validation; large language models

1. Introduction

Psychological modeling is a core means of understanding, predicting, and intervening in human behavior, because psychological processes are an endogenous driver of the actions people take [1]. It builds a bridge between psychological theory and computable frameworks by turning latent constructs such as emotion, cognition, and motivation into quantities that can be represented, simulated, and forecast [2,3]. As an interdisciplinary field it has evolved from theory-driven rule-based modeling, through data-driven machine learning, to a paradigm shaped by artificial intelligence (AI) and large language models (LLMs) [4,5]. Its scope has widened in parallel, from individual traits and emotional fluctuations to group-level phenomena such as consensus and behavioral emergence, and further to societal processes such as environmental adaptation and cultural shaping.

Three shifts characterize this trajectory: from theory-driven to data-driven inference, from static characterization to dynamic tracking, and from single-modality description to multimodal fusion [6]. They have delivered clear gains in predictive accuracy, particularly through spatiotemporal deep learning and, more recently, LLMs. They have also introduced a persistent tension. High-capacity models often behave as black boxes whose internal logic cannot be mapped onto psychological constructs, which limits both their scientific value and their safe use in high-stakes settings [7,8]. The methodological landscape has fragmented at the same time, so that comparable problems are attacked with probabilistic graphical models in one piece of literature, deep sequence models in another, and agent-based simulations in a third, with little cross-referencing and few shared benchmarks.

System dynamics and dynamical systems. Because this review appears in a Special Issue on system dynamics, we state at the outset how we use two terms that are often conflated. *System dynamics* (SD) denotes the specific modeling framework introduced by Forrester and consolidated by Sterman, in which a system is represented by stocks, flows,

delays, and feedback loops, typically as a set of coupled ordinary differential equations at the aggregate level, and in which causal-loop and stock–flow diagrams are the primary tools for building and communicating the model [9,10]. *Dynamical-systems modeling* is the broader mathematical practice of representing any process by state variables and a law of motion; it includes SD but also individual-level differential-equation models of affect regulation, discrete-state models such as hidden Markov models, early-warning analyses of critical transitions, and agent-based simulations. In psychology and public health, SD proper has been used to map the feedback loops of major depressive disorder [11], to model the hypothalamic–pituitary–adrenal stress axis [12], to represent self-regulation and goal pursuit in organizational research [13], and to support population-health policy [14]; and the comparison of SD with agent-based models has been used to study when aggregate feedback structure suffices and when individual heterogeneity and network structure matter [15]. In the first version of this review the term “system dynamics” was not part of the search strategy, and SD models were not treated as a distinct family. In this revision we added SD and related terms to the search (Section 2), we treat SD as a named model family in the individual-dynamics and group classes (Sections 4.4 and 4.5), and we treat the stock–flow formulation as one of the aggregation operators that connect levels (Section 3.3). Throughout, “dynamical” refers to the broad practice and “system dynamics” to the Forrester–Sterman framework.

Existing reviews address the fragmentation only partially. Surveys of affective computing map the move from unimodal to multimodal analysis [6]; reviews of LLMs in psychology catalogue emerging uses and risks [5,16]; surveys of agent-based modeling with LLMs examine social simulation [17,18]. Each covers a slice of the field in depth, but none offers a cross-scale account that connects individual cognition to group emergence while remaining explicit about method assumptions and evidence quality. Three substantive gaps lie behind that omission. Many models are validated by statistical fit alone and are not constrained by prior psychological theory, so their outputs cannot be read causally [8]. Most paradigms deliver static snapshot predictions, so the long-horizon evolution of individual states, such as stress recovery or cognitive change, and its coupling to group-level emergence are rarely modeled jointly [19,20]. Models are typically trained on culturally narrow data and deployed without adequate privacy and governance safeguards [21,22]. These gaps restrict deployment in exactly the complex, high-risk settings where reliable psychological modeling would matter most.

This review responds with an organizing framework and with four claims that the framework makes visible. The framework (Section 3) places research objects on nested levels of analysis and sorts modeling tasks into five classes that share a single definition, so that model families can be compared within and across classes on assumptions rather than on non-comparable accuracy figures. The four claims are stated here in the form defended in the rest of the paper. Two of them (C2 and C3) are stated more narrowly than in the first version of this review, and all four are given boundary conditions and counterexamples in Section 4.7; the claims are conclusions we draw from the reviewed literature, and we say in each case what observation would count against them.

C1 Models inherit their ontology from a psychological theory, and theories differ systematically in *computational readiness*, which we define as the extent to which a theory can be implemented in a computational model without adding assumptions the theory does not itself state. Scored on the rubric of Section 3.1, that gradient, and not the capability of learning algorithms, currently determines where theory-informed learning (theory embedded as a constraint inside a learned function class) is feasible (Sections 3.1 and 7.2).

- C2 Along the ladder from individual to collective units, the *dominant use* of models shifts from estimation, in which model parameters, whether of a descriptive or a generative model, are fitted to the same units' data and validated psychometrically, to simulation, in which micro-rules are posited or borrowed and validated by reproducing macro patterns. This is a shift in estimand and validation target, not in model form: forward-form models are common at the individual level and inverse inference exists at the group level. Validation practice has not followed that shift, and the psychometric standards of the individual classes have no settled counterpart in generative simulation (Sections 4 and 4.6).
- C3 The binding constraint on cross-scale modeling is multilevel data rather than model capacity. Resources that observe the same people both as individuals and as members of a modeled collective exist, but they are few, short, small, and rarely record the collective outcome that simulations target, so the aggregation operator and the split between individual susceptibility and structural exposure remain unidentified (Sections 4.5 and 5).
- C4 The evidence base is heterogeneous in quantity and methodological quality. Single-dataset accuracy is abundant, external validation is scarce, and randomized evidence of benefit exists, at the time of writing, for one generative intervention, so claims of readiness have to be indexed to an explicit evidence tier rather than to benchmark scores (Sections 2 and 6).

The remainder of the paper is organized accordingly. Section 2 states the review methodology, including the search strings, the selection flow, and the evidence-tier coding used throughout. Section 3 develops the framework, scores the computational readiness of the theories that supply model ontologies, and positions the review against existing surveys. Section 4 compares model families class by class and closes with the boundary conditions and counterexamples for C1–C4. Sections 5 and 6 cover datasets and applications, with military use treated as a bounded and ethically flagged case. Section 7 converts five cross-cutting bottlenecks into research-practice recommendations and, where the material allows it, testable predictions.

2. Review Methodology

This is a structured narrative review rather than a systematic review with quantitative synthesis, because the heterogeneity of methods, tasks, and metrics in this field makes meta-analysis inappropriate. A narrative design does not, however, remove the need to report how the literature was found, selected, and judged. We therefore report the search, the selection flow, the coding scheme, and the evidence-tier rules below, following the reporting logic of PRISMA 2020 and PRISMA-ScR [23,24], so that a reader can see what the corpus does and does not cover and can reproduce or challenge the judgments that rest on it. The full search strings, the list of included studies with their codes, and the criterion-level readiness scores of Section 3.1 are provided as Supplementary Material.

Scope of “psychological” and “behavioral.” A study was in scope when it modeled a psychological construct (a state, trait, capacity, process, or collective property that is defined psychologically) from, or in relation to, an observable. “Behavioral” observables were taken to include overt actions and choices, response times, verbal and written content, nonverbal behavior (facial expression, gaze, posture, prosody), movement and trajectories, interaction patterns (who interacts with whom, when, and how), and physiological signals used as indices of a psychological state. Studies whose observable was modeled without a psychological construct (for example, mobility forecasting driven by geometry alone) were out of scope. Following one reviewer's suggestion, we record for each included study both

the construct and the observable, and the datasets of Section 5 are organized by construct, unit of analysis, and observable rather than by modality alone.

Sources and search strings. We searched Web of Science Core Collection, Scopus, IEEE Xplore, the ACM Digital Library, PubMed, PsycINFO, and arXiv. The original search was run in March 2026 and updated in May 2026; the supplementary search for system-dynamics work described below was run in August 2026. The core string combined a construct block with a method block: (*psycholog** OR *behavio*r** OR *affect** OR *emotion** OR *cogniti** OR *personality* OR *motivation* OR *self-regulation* OR *resilience* OR *group psychology* OR *collective behavio*r**) AND (*computational model** OR *mathematical model** OR *machine learning* OR *deep learning* OR *neural network* OR *agent-based* OR *structural equation* OR *latent variable* OR *hidden Markov* OR *dynamical system** OR *large language model** OR *LLM*). In response to the Special Issue scope, the revised search added a third block for the system-dynamics tradition: (*system dynamics* OR *stock-and-flow* OR *stock and flow* OR *causal loop* OR *feedback loop model**) AND the construct block. Fields searched were title, abstract, and keywords; the language was English; publication years were 2015 to 2026 for the database search, with no lower limit for records admitted by citation tracking (see corpus layers below). Backward and forward citation tracking was applied to every record retained after full-text screening.

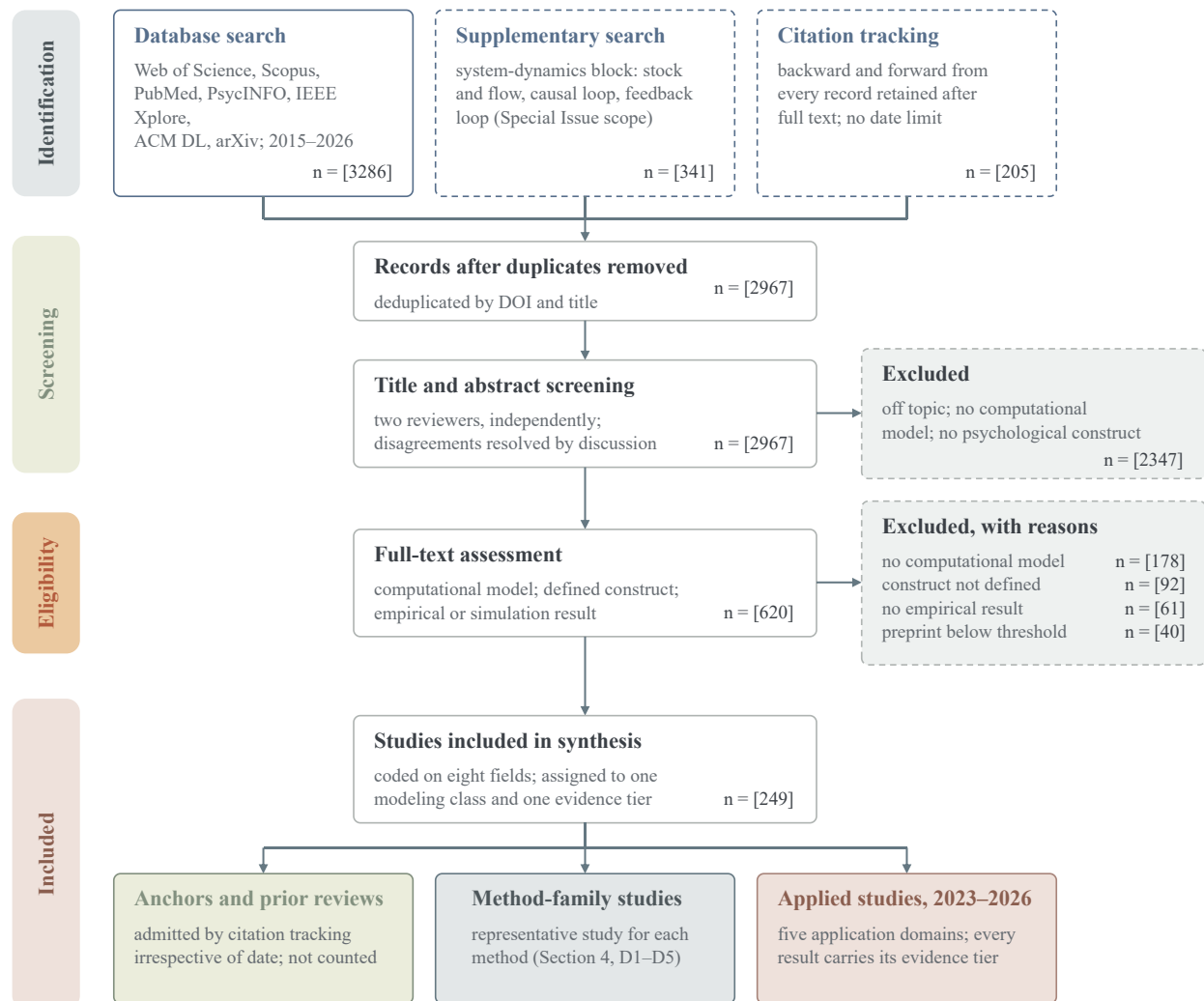
Publication types and preprints. Journal articles, conference papers with archival proceedings, and books or book chapters were eligible. Preprints were eligible only from arXiv, PsyArXiv, and OSF, only when published within the last three years at the time of the search, and only when they (i) reported a computational model with an empirical evaluation and (ii) had either at least 20 citations in Google Scholar at the time of screening or a peer-reviewed successor that we could not yet cite in final form. Every preprint is marked with the superscript “PP” at each citation in the text and tables and is coded at the lowest evidence tier regardless of its citation count. Where a preprint has since appeared in peer-reviewed form we cite the published version.

Screening and selection. Records were deduplicated by DOI and title. Title and abstract screening and full-text screening were each carried out independently by two authors against the criteria that a study (i) proposed or applied a computational model, (ii) targeted a clearly defined psychological or behavioral construct in the sense above, and (iii) reported an empirical evaluation, a validated simulation, or a concrete application. Purely qualitative work, opinion pieces without a modeling component, and studies whose observable carried no psychological content were excluded. Disagreements were resolved by discussion, and unresolved cases were referred to a third author. Figure 1 reports the number of records at each stage and the reasons for exclusion at full text; the counts refer to the search-and-selection process, not to the number of references cited, which is larger because it includes the theoretical anchors and prior reviews described next.

Selection of representative studies within the included set. A narrative review cannot discuss every included study, so we state the rule by which studies appear in the text. For each model family within each modeling class, we cite (a) the study that introduced the family or its standard formulation, (b) the study with the strongest evidence tier for that family (Table 1), and (c) up to three recent applied studies chosen to span different constructs or domains. The remaining included studies are listed, with their codes, in the Supplementary Material. “Canonical empirical milestones” means studies in (a) or (b) that are cited by at least two of the prior reviews.

Coding scheme. Each included study was charted on eight fields: modeling class (D1–D5, by the assignment rule of Section 3.4); construct; observable; unit of analysis and interaction structure (individual, dyad, small group, group, population); model family; data modality; validation type (parameter recovery in simulation, in-sample fit, cross-validation within one dataset, external or out-of-distribution validation, prospective validation, ran-

domized evaluation); and evidence tier. Two authors coded each study; disagreements on class, validation type, or tier were resolved by discussion and, when needed, by a third author. The qualitative fields that feed the comparison tables of Section 4, namely core assumption, data requirement, and characteristic failure mode, were drafted by one author from the coded studies and checked against the source papers by a second.



Bracketed counts are reported in the Supplementary Material with the full search strings and the coded list of included studies.

Figure 1. Selection flow of the review. Records at each stage of identification, screening, eligibility, and inclusion, with the reasons for exclusion at full text. Dashed frames mark the two sources added in this revision, the supplementary system-dynamics search and citation tracking. The included set splits into the method-family studies that populate Section 4 and the applied studies of 2023–2026; theoretical anchors and prior reviews enter by citation tracking irrespective of date and are not counted.

Evidence tiers. Because a recurring criticism of the first version was that preprint, simulation, and in-sample findings were presented at the same level as replicated evidence, every study is assigned one of four tiers (Table 1), and the tier is stated wherever a finding is used to support a claim. The tiers grade the evidence for a finding, not the quality of the study: a well-designed simulation study is tier E2 by construction. Where a section discusses a cluster of studies, we state the highest tier present in the cluster and note when it is E2 or lower.

Table 1. Evidence tiers used throughout the review. The tier of a finding is stated in the text and tables.

Tier	Label	Definition
E1	Provisional	Preprint or unrefereed report, or a peer-reviewed study reporting a model without an empirical evaluation
E2	Single-setting	Peer-reviewed; evaluation by in-sample fit, cross-validation within one dataset, or simulation without calibration against independent real data
E3	Externally validated	Peer-reviewed; evaluation on an independent dataset, population, site, or period, or prospective validation, or independent replication of the main result
E4	Causally identified	Randomized or otherwise design-identified evidence that the model, or an intervention built on it, changes a real-world outcome

Corpus composition. The included studies fall into four layers, and we report the layers because they shape what the review can support. *Theoretical and methodological anchors* were admitted by backward citation tracking irrespective of date, among them prospect theory, the social-force model, the circumplex and appraisal accounts of emotion, the Forrester–Sterman system-dynamics framework, and the transformer architecture, because later work is unintelligible without them. *Canonical empirical milestones*, defined above, supply the load-bearing evidence for claims that recur across sections. A layer of *prior reviews* is used to position this one; we cite reviews to position and to summarize, and we do not use a review as the sole support for an empirical claim. The largest layer comprises *applied modeling studies*, predominantly from 2023 to 2026. This layer is recent and reflects current practice, but it is correspondingly thin on replication: the majority of its entries are tier E2, and several are preprints at tier E1. The composition is itself the first indication of claim C4.

Synthesis. Retained studies were grouped into the five modeling classes of Section 3 by the assignment rule stated there. Because reported accuracies derive from different datasets and protocols, we compare methods qualitatively, by assumptions, interpretability, data requirements, and failure modes, and treat any numeric result as valid only within its original dataset and protocol. Field-wide statements about frequency (“scarce,” “few,” “most”) refer to the coded validation-type and tier fields of the included set, and the corresponding counts are given in the Supplementary Material; where a statement refers to the literature beyond the included set we say so and phrase it as an impression rather than a count.

3. Framework and Theoretical Foundations

Psychological and behavioral modeling transforms internal processes (cognition, emotion, motivation, decision-making) and overt behavior (action, interaction, coordination) into computable, testable, and predictive models. It draws on mathematical modeling, machine learning, systems engineering, and control theory to represent the dynamic coupling between psychology and behavior, and it supports applications in human–AI interaction, decision support, organizational management, and emergency management. Figure 2 previews the review as a whole: the three historical paradigms, the four decisions that define any model in this field, and the structure of the review with the section in which each claim is defended.

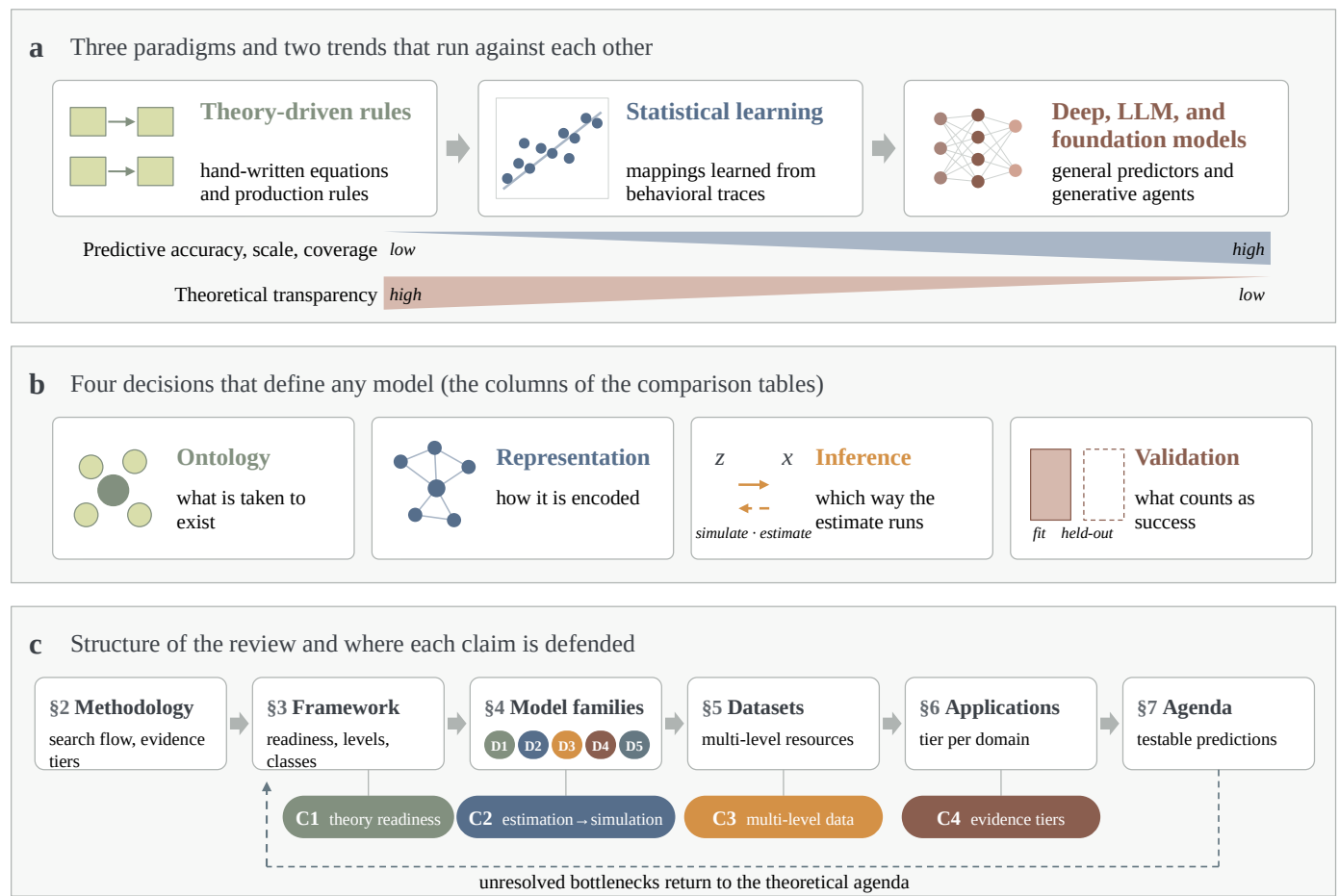


Figure 2. Roadmap of the review. (a) The three paradigms the field has passed through; the two wedges show predictive accuracy and coverage rising from left to right while theoretical transparency falls. (b) The four decisions that define any model and that the comparison tables of Section 4 report as their columns. (c) The sections of the review, the section under which each claim C1–C4 is defended, and the five modeling classes D1–D5 that Section 4 compares; the dashed loop is the path by which unresolved bottlenecks return to the theoretical agenda.

3.1. Psychological Theories as Modeling Substrates

Every computational model in this field inherits its ontology from a psychological theory, whether or not that inheritance is acknowledged. A classifier that outputs one of six emotion labels has adopted basic-emotion theory, and a regressor that outputs valence and arousal has adopted the circumplex. Because such commitments determine what a model can represent and how its failures should be read, we begin with the theory families that recur as modeling substrates and with an assessment of how *computationally ready* each family is. We define computational readiness as the extent to which a theory can be implemented in a computational model without adding assumptions that the theory does not itself state. One reviewer pointed out, correctly, that this cannot be a single scale, because theories are ready in different ways: a trait theory may supply a validated measurement framework and no dynamics, whereas a decision theory may supply a generative process and only a thin measurement tradition. We therefore score readiness on two axes, defined in the rubric below, and we use the first of them, formal specification, to state claim C1. Table 2 reports both scores and the rule by which they combine, and Figure 3 places every family on the two axes.

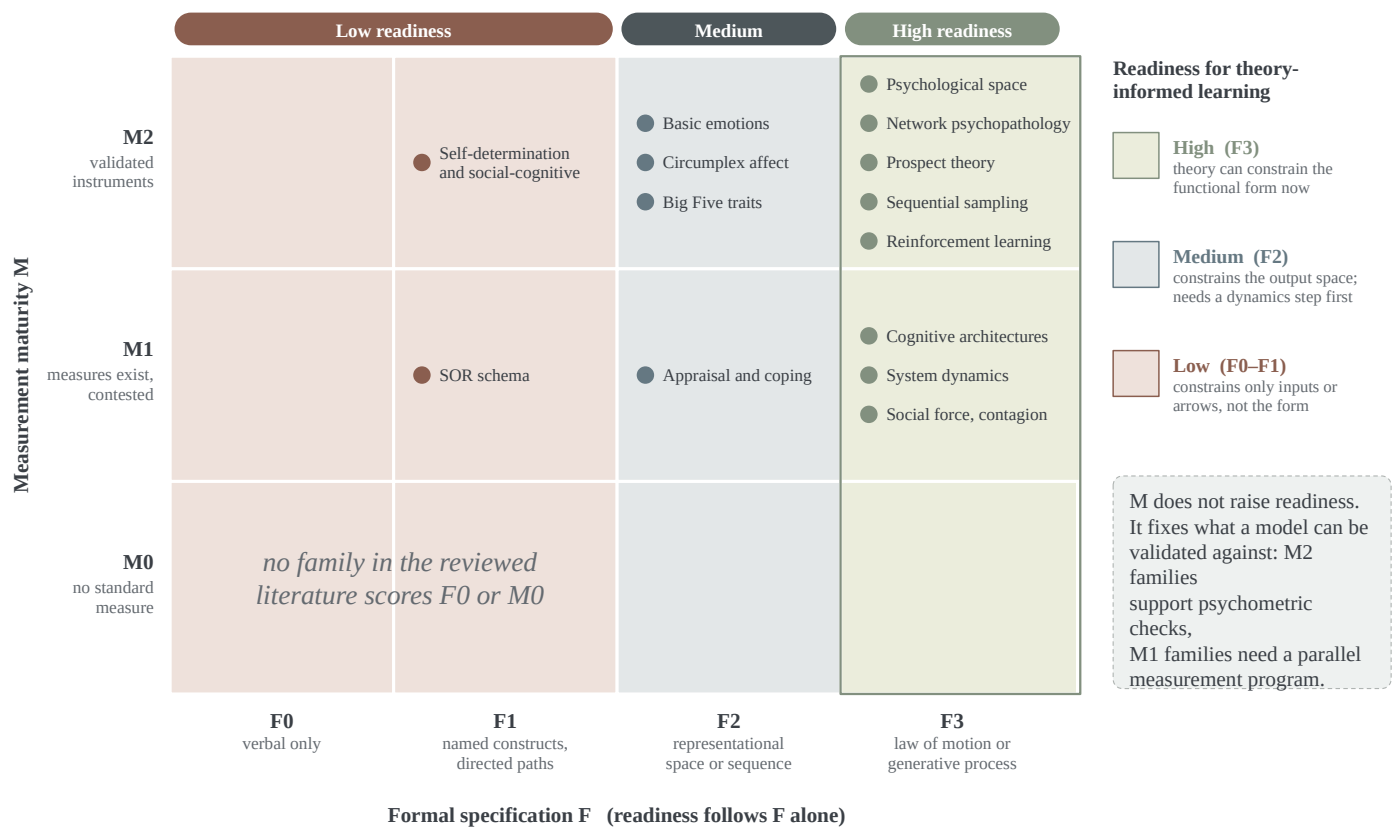


Figure 3. Two axes of computational readiness. The fourteen theory families of Table 2 placed by formal specification F (columns) and measurement maturity M (rows). Readiness follows F alone, so the column bands, not the rows, carry the low–medium–high grading; the outlined F3 column holds the families that can constrain the functional form of a learned model, which is where theory-informed hybrids have appeared first (claim C1). M fixes what a model can be validated against: the M2 families in the upper row support psychometric validation, whereas the M1 families of the F3 column specify a process but need a parallel measurement program. No family scores F0 or M0.

Emotion. Three competing accounts anchor affective modeling. Basic-emotion theory posits a small set of discrete, universal emotion programs [25] and underwrites the categorical labels of most recognition systems, including early artificial-psychology architectures [2]. The circumplex model treats affect instead as positions in a continuous valence–arousal space [26] and underwrites dimensional regression and the annotation schemes of most physiological corpora. Appraisal theory, in Scherer’s component-process form, models emotion as the output of a sequence of evaluations of an event’s relevance, implication, and coping potential [27], with the Lazarus tradition supplying the parallel account for stress [28]. The three are not interchangeable. Categorical and dimensional schemes are computationally cheap but theoretically thin, whereas appraisal structure is expensive to operationalize and is the only account that explains *why* a stimulus produces a given emotion. Interpretable screening frameworks that pair learned features with psychologically meaningful intermediate variables move in this direction [3], although the framework we cite does not itself adopt an appraisal ontology. Affective computing as a research program, that is, machines that recognize, model, and express emotion, was set out by Picard and still frames the sub-area [29].

Personality and individual differences. The lexical tradition consolidates trait structure into the Big Five [30], which supplies the label space for essentially all text-based personality prediction. Its instruments are among the most thoroughly validated in psy-

chology, which is why personality scores high on the measurement axis of Table 2 even though the theory states no dynamics.

Table 2. Psychological theory families as computational modeling substrates, scored on the two-axis rubric. *F*: formal specification (*F0* verbal; *F1* variables and directed relations; *F2* representational space or ordered sequence; *F3* functional form or generative process). *M*: measurement maturity (*M0* none; *M1* partial or contested; *M2* established). Readiness for theory-informed learning is high if *F3*, medium if *F2*, low otherwise. The last column gives the criterion that fixed *F*.

Theory Family	Typical Computational Operationalization	Main Classes	F	M	Readiness	Basis for F
Basic emotions [25]	Categorical emotion classifiers; discrete affect modules	D1	2	2	Medium	Fixed label set; no dynamics
Circumplex affect [26]	Valence–arousal regression; dimensional annotation	D1, D4	2	2	Medium	Metric space; no law of motion
Appraisal and coping [27,28]	Appraisal-variable pipelines; interpretable affect frameworks	D1, D3	2	1	Medium	Ordered check sequence; check functions unspecified
Big Five and lexical traits [30]	Trait label spaces for text and behavior prediction	D1, D2	2	2	Medium	Five-factor space; no process
Similarity and psychological space [31]	Geometric representations estimated from similarity data	D1	3	2	High	Explicit similarity function over a metric space
Network psychopathology [32]	Symptom–interaction network estimation	D1, D4	3	2	High	Pairwise Markov random field or dynamic network with stated update rule
Prospect theory [33,34]	Parametric value and weighting functions in choice models	D3	3	2	High	Parametric value and weighting functions
Sequential sampling [35,36]	Drift–diffusion and preference–accumulation models	D1	3	2	High	Stochastic accumulation process with likelihood
Cognitive architectures [37]	Production-rule simulation of task behavior	D1, D3	3	1	High	Executable production system and memory equations
Reinforcement learning [38]	Reward-driven policy and learning models	D2, D3	3	2	High	Value and policy update rules
System dynamics (stock–flow) [10,11]	Stock–flow–feedback models of aggregate psychological processes	D4, D5	3	1	High	Coupled ordinary differential equations over stocks
Self-determination and social-cognitive [39,40]	SEM path structures; mediation hypotheses	D2	1	2	Low	Named constructs and directed paths; no functional form
SOR schema [41]	Cue → state → behavior path models	D3	1	1	Low	Three-stage path; stages unspecified
Social force and contagion [42]	Interaction-force and compartmental dynamics	D5	3	1	High	Force equations; compartmental rate equations

How constructs are represented and measured. Three commitments cut across theory families and are stated here rather than under any single family, because they apply to the representation of psychological constructs in general. The first is the choice between reflective latent variables, in which a construct is a common cause of its indicators, and the network view, in which a construct is a system of mutually reinforcing components [32,43]; the choice is consequential because it changes what an intervention target is. The second is

the geometric representation of psychological content, in which similarity judgments define a space that models can inherit directly [31]. The third is measurement theory, comprising classical test theory, item-response theory, and their Bayesian extensions, which provides the validity apparatus on which any state- or trait-estimating model implicitly relies [44]. The second axis of the readiness rubric is built on this third commitment.

Judgment, decision, and cognition. Decision theory offers the most formally specified substrates in psychology. Prospect theory specifies a parametric value function and probability weighting [33,34]; sequential-sampling models specify the full generative process for choices and response times [35], with decision field theory extending the logic to dynamic preference [36]; dual-process theory organizes when deliberative rather than intuitive processing dominates [45]; and cognitive architectures such as ACT-R specify the mind at the level of production rules and memory modules [37], with the rational analysis of memory supplying an explicit law for forgetting and retrieval [46]. Reinforcement learning supplies the shared formal language of learning under feedback and is unusual in being simultaneously a psychological theory, a neuroscientific hypothesis, and an engineering toolkit [38]. Executive control and attention have a strong measurement tradition built on latent-variable models of task batteries [47] but a thinner generative one, while their formal models remain few; these constructs enter the review in Section 4.1.

Motivation, self-regulation, and social behavior. Self-determination theory, built on competence, autonomy, and relatedness [39], and social-cognitive theory, built on self-efficacy and observational learning [40], dominate the applied structural-equation literature of Section 4.2, while the stimulus–organism–response schema [41] organizes person–environment work in Section 4.3. At the collective level, the social-force analogy [42] and contagion metaphors ground the mechanistic crowd and opinion models of Section 4.5, and the stock–flow–feedback formulation of system dynamics supplies a further substrate for aggregate psychological processes such as the accumulation and depletion of stress or motivation [10,13].

The consequences of this inheritance are practical, because the ontology fixes what an error means. Consider one interview corpus modeled three ways. Under basic-emotion theory the output is one of a small number of labels, and confusing sadness with fatigue is a category error of the same weight as confusing sadness with anger [25]. Under the circumplex the same confusion is a short displacement in valence–arousal space and is barely penalized, whereas a sign error in valence is severe [26]. Under an appraisal account neither is primarily a labeling error, since the model has misjudged the evaluation that produced the response, and the diagnostic question becomes which appraisal was mis-estimated [27]. The three models can be trained on identical data and report similar aggregate accuracy while disagreeing about what has gone wrong and about what should be repaired. This is the first reason the present review compares model families by assumptions rather than by reported accuracy: accuracy figures computed under different ontologies are not commensurable.

Readiness rubric and rating procedure. Two axes are scored for each theory family. The *formal-specification* axis *F* records how much of the mapping from psychological process to observable the theory states: *F0*, verbal only; *F1*, names the variables and the directed relations among them (a path structure) but no functional form; *F2*, specifies a representational space, label set, or ordered sequence with enough structure to constrain what a model may output, but not how it changes; *F3*, specifies a functional form, update rule, or generative process from which behavior can be simulated. The *measurement-maturity* axis *M* records whether the constructs can be measured in a way that supports validation: *M0*, no standard instrument or paradigm; *M1*, instruments or paradigms exist but psychometric evidence is partial or contested; *M2*, standardized instruments or task paradigms with

established reliability, validity, and, where relevant, invariance evidence. Readiness for theory-informed learning, the quantity used by claim C1, is a function of F alone: *high* if $F3$, *medium* if $F2$, *low* if $F0$ or $F1$. The rule is stated so that it can be disputed: a reader who believes that a path structure is enough to constrain a learned model will assign high readiness at $F1$ and will disagree with C1 at exactly the point where the disagreement lies. M does not enter the readiness rule but governs how the resulting model can be validated, which is why it is reported. Two authors scored the fourteen families independently against the criteria; initial disagreements occurred on 3 of 28 scores, all at the boundary between adjacent formal-specification levels, and were resolved by discussion, with the criterion-level justifications recorded in the Supplementary Material. The scores are rubric-based judgments, not measurements from a validated instrument, and C1 is qualified accordingly in Section 4.7.

The rubric makes the pattern in Table 2 and Figure 3 explicit. Prospect theory, sequential sampling, and reinforcement learning are $F3$ and can constrain a learned model directly. Basic-emotion, circumplex, and Big Five theories are representational commitments ($F2$) that constrain label spaces but not dynamics, even though their measurement tradition is at least as strong. Self-determination and social-cognitive theories are $F1$: they name constructs and directed paths, and their operationalization in path models is a loose translation in which the theory constrains arrows but not functional form. Network psychopathology is $F3$ because the estimated object is a Markov random field or a dynamic network with a stated update rule, which is what distinguishes it from the $F1$ path theories despite the superficial resemblance of both to a graph. This gradient on F , rather than any property of learning algorithms, is the main reason that theory-informed hybrids have appeared first in memory and decision modeling [48] and remain rare in motivation and social psychology. The gradient on M explains a second pattern: the classes with the strongest validation practice ($D1$, $D2$) are those whose substrates are $M2$.

A second consequence concerns validity. Every model in the individual classes is estimated against an instrument, usually a scale or an annotated label, and the instrument is not the construct. Measurement theory supplies the apparatus that keeps the two apart through construct validity, reliability, and measurement invariance across groups and occasions [44]. Two results show how much this still matters in computational practice. Tasks with large and replicable group-level effects can have poor individual-level reliability, so a model fitted to individual differences may be fitting noise unless estimation is designed for that regime [49]. Self-report instruments are also open to strategic distortion, which forced-choice designs and formal faking models were developed to detect [50]. A model that faithfully reproduces a distorted or unreliable score has succeeded at the wrong task, and no interpretability method applied after the fact repairs that.

3.2. Anatomy of a Model: Four Decisions the Review Compares

Methods in this field arrive from psychometrics, cognitive science, statistical physics, systems engineering, and machine learning, and they are rarely described in a common vocabulary. To compare them at all, we describe any model by four choices, shown in panel (b) of Figure 2.

Ontology. What is taken to exist and therefore to be estimated: six discrete emotion programs, two affective dimensions, a set of appraisal checks, a symptom network, a drift rate, a stock of accumulated stress. This choice is inherited from theory in the sense of Section 3.1, and it is the choice most often left implicit.

Representation. How the ontology is encoded in a computable object: a reflective latent variable, an undirected network, a set of production rules, a differential equation, a stock–flow structure with feedback loops, a geometric space, or a natural-language persona.

Two models may share an ontology and differ entirely here, as categorical affect does when it is encoded as a softmax output rather than as a discrete state in a Markov chain.

Inference. Which way the estimate runs. Inverse inference recovers a latent quantity from observed behavior; forward inference instantiates a posited quantity and generates behavior from it. The two are not exclusive: a generative model such as the drift-diffusion model is forward in form but is used inversely when its parameters are fitted to a person's choices. What varies across the five classes is the dominant use, which is the subject of claim C2.

Validation. What counts as success: recovery of known parameters in simulation, construct validity against an established instrument, out-of-sample prediction, or calibration against real collective data. Validation is where the field's practice is least standardized. Section 4.6 argues that the standard actually applied in a class tends to be the one inherited with the dominant use, rather than the most severe test the available data would permit; "severe" here means a test that the model could fail and that would distinguish it from alternatives.

Reading the historical paradigms through these four decisions clarifies what changed at each step. Rule-based modeling fixed ontology and representation by hand and validated by qualitative correspondence to known phenomena. Statistical learning kept the ontology, typically a questionnaire score or an annotated label, but replaced the hand-specified representation with a learned one, which is why the accuracy gains were immediate and the interpretability loss was equally immediate. Theory-informed learning re-imposes theory at the representation layer while keeping a learned function class [51,52]. Foundation models of cognition go further and blur the first decision, because the ontology is no longer a small set of declared constructs but whatever the training corpus induces [53], which is why the dispute they provoke is about explanation rather than about fit [54,55]. System dynamics sits at the opposite end of the same spectrum: ontology and representation are fixed by hand as stocks and loops, and validation rests on reproducing reference modes and on structural tests of the loops [10]. The comparison tables of Section 4 report the first, second, and fourth of these decisions as their assumption, data-requirement, and failure-mode columns, and the third organizes the classes themselves.

3.3. Cross-Scale Research Objects

The field's research objects sit on a ladder of nested levels that differ in their characteristic constructs, observables, and timescales (Figure 4). One reviewer asked us to separate constructs from observables and to state which levels the review actually covers in depth; we do both here. At the *individual* level the constructs are states (emotion, stress), traits (personality, values), capabilities (resilience, self-regulation), and processes (attention, memory, decision); the observables range from millisecond physiological signals and response times through language, facial and vocal expression, and smartphone traces to multi-year survey waves. At the *dyadic and small-group* level, which the first version of this review did not treat separately, the constructs are interpersonal and interaction-borne: rapport, synchrony, trust, leadership emergence, and shared affect; the observables are turn-taking, joint physiology, gaze and posture coordination, and communication content. At the *group* level the constructs are cohesion, consensus, contagion, and collective risk perception, whose defining property is that they are not attributes of any single member; the observables are interaction structure, message cascades, and joint trajectories, and coordinated movement is one such observable rather than a construct. The *societal* level concerns slowly varying distributions of values, trust, well-being, and norms observed through repeated surveys and large-scale digital traces. This review covers the individual and group levels in depth and treats the societal level as an extension of the group level: at

that resolution the modeling targets are the same collective properties observed through sparser instruments, and the literature that models them with the methods reviewed here is thin. The framework is therefore best read as a two-level framework, individual and group, with the dyad as the smallest collective and the society as the largest.

Why the levels must be linked. Single-level models are partial, but that is not by itself a reason to link levels. The reason is that several questions cannot be posed, let alone answered, without data at more than one level. First, whether a collective outcome is driven by individual susceptibility or by structural exposure: emotional contagion in a team can arise because some members are easily influenced or because of who talks to whom, and the two explanations have different intervention targets but the same aggregate signature. Second, whether an intervention that changes individual states changes the collective outcome: a stress-reduction program validated on individual scores may or may not change team performance, and the cross-level effect is the quantity of interest. Third, which aggregation operator holds: mean-field, network-mediated, agent-based, and stock–flow operators (below) make the same macro predictions in many settings and differ only in what they predict when the interaction structure or the heterogeneity of members changes. Fourth, the direction and timescale of the downward path, that is, how quickly norms and feedback constrain individual behavior. Each of these is a question about a cross-level parameter, and Section 5 states what would have to be measured, and at what resolution, to identify it.

Two properties of the ladder shape everything that follows. The levels are coupled in both directions, since individual states aggregate into collective phenomena while group and societal contexts constrain individual behavior through norms and feedback. The levels also differ systematically in measurement resolution, with dense signals at the bottom and sparse aggregates at the top, which is why the micro–macro linkage identified in Section 4.6 as the field’s binding data constraint cannot be solved at the modeling layer alone.

Aggregation operators. Coupling between levels is itself a modeling choice with testable assumptions, and five operators account for most of the literature. Mean-field and compartmental formulations replace the population with densities and assume that any two members interact with equal probability [56]. Stock–flow formulations, the system-dynamics operator, likewise work at the aggregate level but represent the accumulation and depletion of a quantity through explicit feedback loops and delays rather than through contact rates, and they are the natural choice when the collective process is a slow accumulation, such as the build-up of burnout or the erosion of trust, rather than a contagion [10,14]. Network-mediated formulations retain who is connected to whom and let structure carry the aggregation, an assumption that large field experiments on behavioral spread have tested directly [57]. Agent-based simulation abandons closed form and aggregates by execution, which admits heterogeneous members at the cost of analytic transparency [58]. Statistical multilevel models are the most conservative option, since they do not derive the macro level at all but estimate how much variance sits between groups [59]. Direct comparisons of these operators are rare but informative: the same diffusion process modeled as a stock–flow system and as an agent-based network model gives matching aggregates when the network is well mixed and diverging ones when it is clustered or heterogeneous [15], which is a test of the operator rather than of the psychology. Applied work rarely compares its operator against the alternatives, yet the operator determines whether an emergent pattern has been explained by interaction structure or merely reproduced by a well-chosen density.

Timescale matching. A process can be recovered only if it is observed faster than it moves. Early-warning indicators of state transitions require sampling faster than the dynamics they track [60], which is why ecological momentary assessment rather than

annual survey waves is the enabling instrument for the dynamic class [61], and why between-person cross-sections cannot substitute for within-person processes however large the sample [62]. Timescale mismatch is the most common way in which a study at one level fails to speak to the level it claims to address.

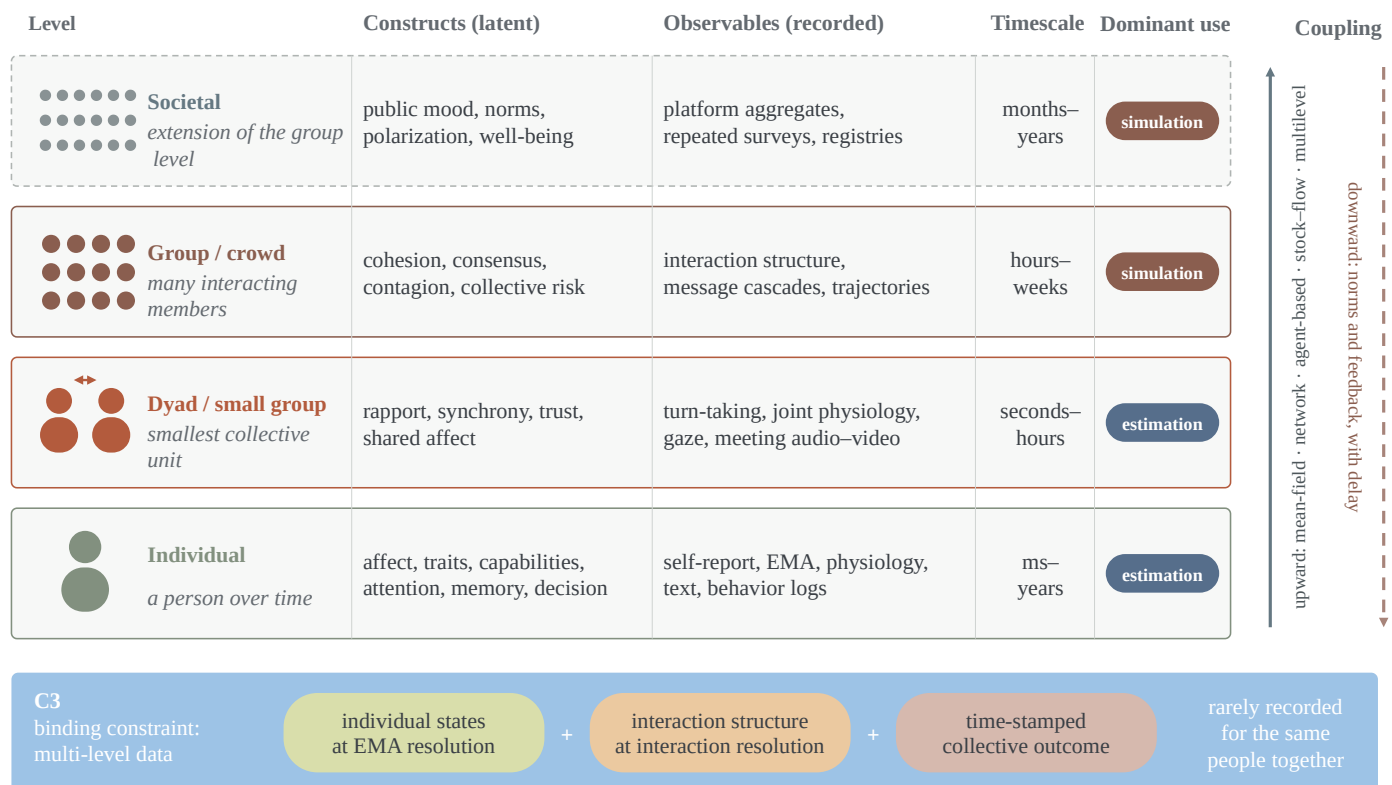


Figure 4. Ladder of levels of analysis. Rows are the nested levels, with the dyad as the smallest collective and the societal level (dashed) as an extension of the group level rather than a third level reviewed in depth. The columns separate the constructs that are modeled from the observables that are recorded, and give the characteristic timescale and the dominant use of models, which changes from estimation to simulation between the dyadic and the group rows (claim C2). The two arrows at the right are the coupling directions: the upward path is carried by one of the five aggregation operators of Section 3.3, the downward path by norms and feedback. The band at the foot states the three components that a multilevel resource must record for the same people, which is the constraint of claim C3 (Section 5).

3.4. Five Modeling Classes: Definition, Assignment Rules, and Scope

To organize methods without forcing them into a single disciplinary silo, we characterize every modeling task by the same triple: the *unit of analysis* that carries the modeled property, the *target* that the model estimates or generates, and the *direction of inference* that relates target to observation. Applying the triple yields five classes. For an individual unit, the target is either the current value of a psychological variable (D1, individual psychological states), the capacity that governs behavior under demand rather than at rest (D2, individual behavioral capabilities), the coupling between the person and a specified context (D3, individual–environment interaction), or the law of motion that governs how the variable changes (D4, individual dynamic change). For a collective unit, the target is a property that no member possesses individually (D5, group psychology). The first version of this review called these five “dimensions.” A reviewer observed that they are not independent axes along which a study can be located but categories into which a study is sorted, and that the scheme was more plausibly defined first and described by the triple afterwards than derived from it. We accept both points. We now call them modeling classes,

we present the triple as the assignment rule rather than as a derivation, and we withdraw the claim that the classes are jointly exhaustive: they cover the included literature, and we have not tested whether independent raters would assign boundary cases to them with adequate agreement, which is the test that a claim of exhaustiveness would require.

Assignment rules for boundary cases. Several overlaps noted by the reviewers are real, and the following rules state how we resolved them. States and traits both fall in D1 because D1 is defined by the estimand, a person-level value at one time, not by its stability; the state–trait distinction is a difference of timescale that D4 handles by asking whether a law of motion is the target. Resilience and self-regulation are assigned to D2 rather than D1 when the estimand is a response to demand, that is, when the model uses behavior under load, over repeated occasions, or after a perturbation; when the same constructs are scored from a questionnaire at rest they are D1 estimands. Person–environment interaction is present in some form in almost every model, so D3 is assigned only when the context is a modeled variable with its own parameters, such as a cue whose effect on state is estimated, an environment with a stated transition function, or an exposure with a dose–response form, and not when context enters as a covariate. Dyadic and interpersonal processes are assigned to D5 with the dyad as the smallest collective, because the estimand, such as synchrony or rapport, is a property of the pair rather than of either member; where a study estimates one partner’s state from the interaction, it is D1 or D4. A model that reports affect at time t belongs to D1 even when fitted to a time series, whereas a model whose parameters describe how affect at t produces affect at $t + 1$ belongs to D4. A persona-conditioned agent is assigned to D2 when the object is the enacted capability of one simulated person and to D5 when the object is the pattern generated by a population of them.

Two consequences of the definition matter for the rest of the review. First, classes are fixed by the target rather than by the algorithm, which is why the same technique, whether hidden Markov models, structural equation modeling, agent-based modeling, system dynamics, or reinforcement learning, legitimately appears in more than one class. This target-first organization complements rather than replaces the algorithm-first taxonomies common in affective computing and computational social science [6,18], and it makes the recurring cross-class role of a small set of model families visible (Figure 5). Second, the third element of the triple, the direction of inference, is the subject of claim C2, which Section 4 develops class by class.

Table 3 states the resulting definition for each class. Read as a whole it shows the pattern that claim C2 names in its revised form: the unit of analysis changes once, between D4 and D5, and with it the canonical estimand changes from a quantity attached to a person and validated against that person’s data to a set of rules attached to a population and validated against a macro pattern. The direction column records the dominant use, not the only one: inverse inference at the collective level (estimating cohesion, collective affect, or network parameters from interaction traces) and forward-form models at the individual level (drift-diffusion, reinforcement learning, cognitive architectures) both exist and are listed.

The framework also carries a consequence that could turn out to be false, which we state so that it can be checked. If classes are the right unit of organization, then methods should transfer more readily within a class than within an algorithm family: a laboratory built to estimate one latent capacity should adapt faster to a second capacity than to a group-level simulation task, even when the second task uses the same architecture. Where that fails, the target-first organization is the wrong cut and an algorithm-first taxonomy would serve better.

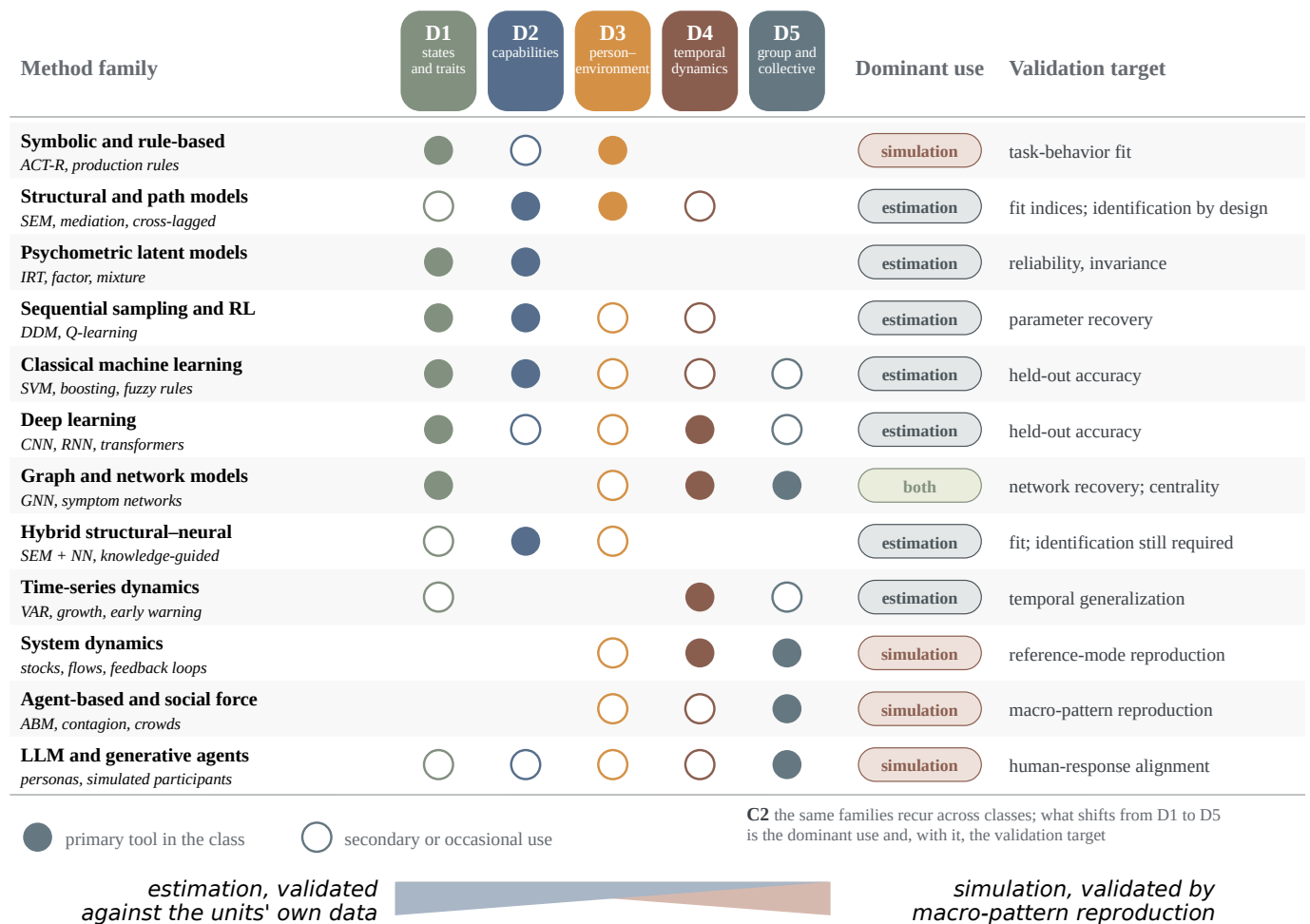


Figure 5. Method families across the five modeling classes. Filled markers mark the classes in which a family is a primary tool, open markers a secondary or occasional use. Several families recur across four or five classes, which is why a target-first organization is more informative than an algorithm-first one. System dynamics is listed as its own family. The two right-hand columns give the dominant use of each family and the validation target that comes with it; the wedge at the foot shows the shift from estimation validated against the units' own data (D1) to simulation validated by macro-pattern reproduction (D5), which is claim C2 in its revised form.

Table 3. The five modeling classes defined by unit of analysis, modeling target, and direction of inference. The direction column gives the dominant use and, in parentheses, the secondary use that also occurs in the included literature.

Class	Unit	Modeling Target	Dominant Use (Secondary)	Canonical Estimand
D1 States	Person	Current value of a psychological variable, whether state or trait	Estimation (forward-form process models fitted to the person's data)	Posterior over a latent value given behavior at one time
D2 Capabilities	Person	Capacity that governs behavior under demand rather than at rest	Estimation (generative agents that enact a capability)	Capability parameter or path coefficients linking antecedents to conduct
D3 Interaction	Person in a modeled context	Coupling between environmental input and psychological response	Estimation (simulation of the person–environment loop)	Response function from cue to state to behavior, or a learned policy

Table 3. *Cont.*

Class	Unit	Modeling Target	Dominant Use (Secondary)	Canonical Estimand
D4 Dynamics	Person over time	Law of motion of the psychological variable	Estimation (forward simulation of stock–flow or dynamical models)	Transition parameters such as autoregression, drift, or dissipation rate
D5 Group	Dyad, group, or population	Property that no individual member possesses	Simulation (estimation of cohesion, collective affect, or network parameters from traces)	Micro-rule or loop parameters that reproduce an observed macro pattern

3.5. Scope and Exclusions

Three boundaries keep the framework from expanding into neighboring fields, and stating them makes the review’s coverage checkable. We include a study only where a psychological or behavioral construct is explicitly modeled, so neural decoding that predicts a stimulus class without a psychological construct, sentiment classification treated purely as a text-labeling benchmark, and clinical trials that report an outcome without a model all fall outside the scope even though each borders it. We exclude modeling of behavior that carries no psychological content, such as mobility forecasting driven by geometry alone. Two adjacent fields, computational psychiatry and computational social science, overlap the framework. A reviewer rightly noted that a field’s organizing principle, by disorder or by data source, is not a reason to exclude its studies. Studies from both fields were screened against the same inclusion criteria as every other record and are cited under the class whose target their result addresses; the point of the earlier remark was only that we do not adopt those fields’ own headings.

3.6. Positioning Relative to Existing Reviews

Table 4 situates this review against the surveys nearest to it and states the contribution as what each of them does not attempt. Affective-computing reviews map recognition methods and the transition from unimodal to multimodal analysis in depth but stay within D1 [6,63]. Mental-health reviews chart machine learning across clinical tasks but are organized by disorder and application rather than by modeling target [64–66]. The rapidly growing literature on LLMs in psychology catalogues instrument-style uses and risks [5,16], and the LLM-agent literature surveys generative social simulation [17,18]; both are method-first and largely silent on the psychometric traditions that govern the individual classes. Collective-behavior reviews connect cognition to crowd simulation but do not treat individual-level inference [67], system-dynamics reviews in public health treat feedback structure but not individual-level measurement [14], and bibliometric mappings quantify the field’s growth and thematic structure without comparing methods [68]. What none provides, and what claims C1 to C4 supply, is a single account that spans the individual-to-group ladder, compares model families by assumptions and failure modes across that span, and states explicitly where the evidence is thin.

Table 4. Positioning of this review against representative prior surveys.

Prior Survey	Scope	What This Review Adds
Affective computing [6,63]	D1 affect recognition; modality fusion	Places affect within a five-class, cross-scale account; adds the theory-readiness analysis of C1
Mental-health machine learning [64–66]	Clinical prediction tasks	Target-first taxonomy across constructs; assumption-level method comparison; explicit evidence tiers (C4)
LLMs in psychology [5,16]	LLMs as instruments and assessors	Integrates LLM roles into each class and applies psychometric validation standards to them

Table 4. Cont.

Prior Survey	Scope	What This Review Adds
LLM agents and computational social science [17,18]	Generative social simulation	Links simulation to individual-level measurement traditions; multilevel validation critique (C2)
Collective behavior modeling [67]	Crowd and group simulation	Connects group emergence to individual state and capability inference; identifies the data constraint (C3)
System dynamics in health and organizations [13,14]	Feedback structure of aggregate processes	Places stock–flow modeling among the aggregation operators and relates it to individual-level measurement
Bibliometric mapping [68]	Publication landscape and topic structure	Methodological judgment rather than description: assumptions, boundary conditions, research agenda

4. Modeling Methods Across the Five Classes

This section traverses the five classes defined in Section 3.4. Each subsection follows the same schema, stating what is being inferred, which model families are used, what each family assumes and delivers, and how it characteristically fails, so that methods remain comparable within and across classes. Because reported accuracies come from different corpora and protocols, families are compared by assumption structure rather than by headline numbers, and each class closes with a comparison table in which the last column gives representative work and the highest evidence tier (Table 1) reached by that family in the included set.

Two observations recur and are stated in advance because they organize the tables. First, a small number of model families act as cross-class workhorses. Latent-variable models (structural equation models, latent class, latent growth), state-space and dynamical models (hidden Markov models, differential-equation and stock–flow models), and, increasingly, LLM-based agents appear in four or five classes each, which is why an algorithm-first taxonomy fragments the field while a target-first one does not. Second, the dominant use of models shifts along the classes, which is claim C2 in its revised form. In D1, D2, and D4 a model, whether descriptive ($P(z | x)$ estimated directly) or generative (a forward model $P(x | z)$ whose parameters are fitted by inversion), is estimated from the data of the units it describes and validated against those units' further data or against an instrument. In D5, and increasingly in D3, the micro-rules that generate behavior are posited or borrowed rather than estimated from the units' own data, and the model is validated by whether the pattern it generates resembles an observed macro pattern. The shift is not from forward models to inverse ones: drift-diffusion, reinforcement-learning, and cognitive-architecture models are forward in form and dominate parts of D1, and inverse estimation of cohesion, collective affect, and network parameters from interaction traces is common in D5. The shift is in what is estimated and against what it is validated, and it is this, rather than any difference in algorithmic sophistication, that explains why validation practice differs so sharply between the individual and the collective classes.

4.1. D1: Individual Psychological States

D1 estimates the components of internal state, namely cognition, emotion, motivation, and personality, from observable signals, and separates momentary fluctuation from trait-level structure. Formally it is the inverse problem $P(z | x)$, and its central methodological question is how much psychological structure to impose on P before fitting.

Measurement-theoretic and latent-variable models. The oldest and most transparent family treats z as a latent variable with an explicit measurement model. Classical test theory, item-response theory, and factor-analytic structures formalize how items relate to constructs and remain the reference standard for construct validity [44]. Bayesian latent-variable formulations extend this to hierarchical and small-sample settings [69], and Bayesian networks

have been used to represent the joint influence structure among developmental factors [70]. The comparative advantage is interpretability and uncertainty quantification; the cost is that the structure must be specified in advance, so mis-specification propagates silently into estimates. A recurring debate concerns whether psychological attributes are best represented as reflective latent causes at all, since the network tradition models constructs instead as systems of mutually reinforcing symptoms [32], a program whose first decade has been systematically reviewed [43] and whose estimation problems in high dimensions remain an active methodological topic [71]. Comparative work on psychodynamic and trait-based structures shows that more than one latent structure can be defended for the same construct domain [72], and the methodological literature explains why fit statistics alone cannot adjudicate between such alternatives [73]. Recent work using generative models to attack the reliability paradox, that is, the observation that tasks with excellent group-level effects can have poor individual-level reliability, shows how much estimation practice still matters at this layer [49].

Process models of cognition. Where the target is a cognitive mechanism rather than a score, sequential-sampling models decompose observed choices and response times into interpretable parameters such as drift rate and decision threshold [35], decision field theory extends the same accumulation logic to dynamic preference formation [36], and cognitive architectures such as ACT-R model task behavior at the level of production rules and memory modules [37]. Hierarchical drift-diffusion formulations have made process modeling tractable for the modest sample sizes typical of social and personality research [74]. These models are forward in form, since they generate choices and response times from posited processes, and inverse in use, since their parameters are fitted to each person's data; they are the clearest case of the transparency dividend, because the parameters are psychologically meaningful by construction, which is why the family is used to test theory rather than merely to predict, provided that the discipline of formal model comparison and falsification is actually applied [75,76]. Mechanism-level modeling has also been extended to unstructured elicitation data such as free lists through Bayesian regression, which brings cross-cultural settings within reach where standardized instruments may not transfer [77]. Mental-control models developed for applied settings such as driving illustrate the same logic in domain-specific form [78].

Cognitive processes beyond affect and personality. A reviewer noted that the first version of this review concentrated on emotion, personality, and mental health and said little about attention and engagement, executive control, memory, cognitive ability, and social cognition. These constructs were within the search scope; the imbalance reflected the density of the applied literature retrieved rather than a scoping decision, and we now state where each sits. Memory is the construct with the most complete formal substrate, since the rational analysis of memory specifies retrieval probability as a function of recency and frequency [46], and it is the substrate on which the first psychology-informed neural network was built [48]. Executive control is organized by a latent-variable structure of inhibition, updating, and shifting estimated from task batteries [47], which supplies an M2 measurement framework for D1 but, as the readiness rubric records, little that constrains a learned model's functional form. Attention and engagement are modeled most often in learning settings, where affective and attentional states are tracked over time [79] and recognized from facial behavior with supervised learning [80] (E2). General cognitive ability has been predicted from resting-state brain networks with out-of-sample validation [81] (E3), a result that belongs to D1 by its estimand although its observable is neural. Social cognition enters D1 in two ways: theory of mind is measured through validated task paradigms, and those paradigms have recently been applied both to humans and to LLMs, with the models matching or exceeding human performance on some tasks and failing

systematically on others [82,83] (E3, replicated across laboratories). Interpersonal preference is instructive as a negative result: machine learning models trained on extensive self-report predicted individuals' general tendencies to desire and be desired but did not predict who would desire whom, in a study with held-out validation [84] (E3); it is a reminder that predictive coverage from traces (below) has limits that a construct's measurement maturity does not remove.

Large-scale behavioral prediction and foundation models of cognition. A distinct recent development sits between process models and black-box prediction. Trained on very large choice corpora, machine-learned value functions have outperformed every classical variant of prospect theory at predicting risky choice while remaining simple enough to inspect, which is an existence proof that theory discovery can be data-driven [85] (E3). The same logic now operates at foundation-model scale. Centaur, a language model fine-tuned on Psych-101, which comprises trial-by-trial data from 160 experiments, 60,092 participants, and more than ten million choices, predicts held-out human behavior better than domain-specific cognitive models across almost every task tested, and generalizes to modified tasks and new domains [53] (E3). Whether such systems *explain* anything is the field's live dispute, and the prediction–explanation distinction [54], the argument that quantitative fit is a poor arbiter among theories [73], and the insistence that computational modeling should force theory building rather than replace it [55,86] all apply with renewed force. Our reading is that foundation models of cognition now set the predictive ceiling against which mechanistic models must be measured, but do not yet discharge the explanatory duties that motivated D1 modeling in the first place.

Statistical learning from behavioral traces. A large applied literature relaxes structural assumptions and predicts psychological state directly from behavioral or demographic features. The founding results showed that ordinary digital records support surprisingly strong trait inference, since Facebook likes predict personality, orientation, and substance use [87] and computer judgments of personality from such traces can exceed those of human acquaintances [88] (both E3, with held-out or cross-sample validation). In the clinical direction, social-media language predicts depression ahead of medical documentation [89,90], a literature consolidated in integrative reviews [91]; the ambient-sensing generalization of this program is digital phenotyping and personal sensing [92,93], with ecological momentary assessment supplying its measurement backbone [61]. Later applied studies extend the recipe across populations and targets, and the cluster is almost entirely E2, that is, single-dataset cross-validated performance: personality from social-media behavior [94], self-rated health [95], burnout risk profiles [96], medication adherence [97], disease-related psychological states from multimodal and genomic features [98], resilience in specific populations [99], and distress predicted from personality with neural networks [100]. Fuzzy-neural hybrids have been applied to psychological-health assessment in specific populations [101], and explainable variants have been used to link childhood maltreatment and positive traits to post-traumatic symptoms [102]. What this family delivers is predictive coverage where no measurement model exists; what it forfeits is construct validity. A model that predicts a questionnaire score accurately has not thereby modeled the construct the questionnaire was designed to measure, a distinction the applied literature frequently elides.

Affect recognition and its modality structure. Emotion is the most methodologically mature sub-area of D1 and the clearest illustration of the paradigm ladder. Its label spaces are inherited directly from the emotion theories of Section 3.1, since categorical systems adopt basic-emotion theory and dimensional systems adopt the circumplex [25,26], and the sub-area as a whole descends from Picard's program for affective computing [29]. Early systems used hidden Markov models over facial and acoustic features and established the state-sequence formulation that later neural models inherited [103,104], and the first-

generation methods were consolidated in an interdisciplinary review [63]. The field then moved decisively to multimodal fusion, and the canonical reviews of that transition document both the accuracy gains and the alignment, representation, and co-learning problems it introduced [6,105], while conversational emotion recognition added interlocutor context as a further structural variable [106]. Physiological signals are a standard observable in this sub-area: electroencephalography, electrocardiography, electrodermal activity, and wearable sensor streams are used to estimate affect and stress in the corpora of Section 5, and the same signals are used at the dyadic level to estimate physiological synchrony (Section 4.5). Contemporary systems apply deep and attention-based architectures to physiological, facial, and linguistic streams, including in interactive settings such as human–robot collaboration [107] and in frameworks that pair learned features with psychologically interpretable intermediate variables [3] (E2). Deep models have also been used at scale to characterize population-level affective responses to events such as the COVID-19 pandemic [108] and to analyze cyber-psychological problems in youth populations [109] (E2). The persistent failure modes, namely modality misalignment, missing modalities, noise sensitivity, and weak ecological validity, now define the sub-area’s research agenda rather than appearing as incidental caveats.

LLMs as instruments and as subjects. Language models ent D1 in two distinct roles that are easily conflated. As *instruments*, they serve as content-analysis tools and replace hand-coded text analysis with learned representations [110], with broad prospects catalogued in recent overviews [5]. As *subjects*, they are probed with the field’s own methods, so that cognitive-psychology batteries characterize their decision and reasoning profiles [111], psychometric inventories elicit apparent trait structure [112,113], and opinion surveys reveal whose views their defaults reflect [114]. The second use requires particular care. Questionnaire responses elicited from an LLM reflect prompt framing and training distribution at least as much as any stable internal property [22], and LLMs do not reproduce the characteristic *response biases* of human survey-takers, responding to question perturbations in ways humans do not [115]. A related and underappreciated problem is response distortion, for which forced-choice designs and formal faking models developed for human assessment are directly relevant to judging whether an elicited profile is genuine or strategically shaped [50]. Table 5 summarizes the existing modeling methods within this dimension.

Table 5. D1—Individual psychological states: model families compared by assumptions, data requirements, and characteristic failure modes, with representative work and the highest evidence tier reached by the family in the included set.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
Latent-variable/measurement models	Construct is a reflective latent cause of item responses	Moderate; validated instrument	Structural mis-specification propagates silently	[44,69,72] (E3)
Psychological networks	Construct is a system of interacting components, not a common cause	Moderate–large; many indicators	Estimation unstable in high dimensions	[32,71] (E3)
Process models (drift-diffusion, ACT-R)	Choice arises from a stated generative process; forward in form, fitted by inversion	Trial-level RT and accuracy	Requires controlled task; poor fit to naturalistic behavior	[35,74] (E3)
Statistical learning from traces	Behavior is predictive of construct without stated mechanism	Large labeled corpora	Predicts the instrument, not the construct	[87–89] (E3)

Table 5. Cont.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
Multimodal deep affect models	Affect is recoverable from fused sensory and physiological streams	Large, aligned multimodal corpora	Modality misalignment; weak ecological validity	[3,6,105] (E2)
Foundation models of cognition	Massed multi-task behavior induces a general predictor	Very large trial-level corpora	Predictive ceiling without explanatory content	[53,85] (E3)
LLM-based inference	Language encodes psychological state; models can be probed	Prompt design; validation set	Prompt and training artifacts mistaken for traits	[22,111,113] (E2)

4.2. D2: Individual Behavioral Capabilities

D2 infers dispositions and competencies such as resilience, self-regulation, adaptive capacity, and decision style from behavior. It differs from D1 in that the target is a capacity to act rather than a current state, which changes what counts as evidence: capabilities are revealed only under demand, so modeling depends on observing behavior under load or over repeated occasions. Table 6 summarizes the existing modeling methods within this dimension.

Structural and mediation models. The dominant approach in applied psychology specifies a directed structure linking antecedents, mediators, and behavioral outcomes, estimated by structural equation modeling or regression-based mediation, with the hypothesized paths usually supplied by self-determination or social-cognitive theory [39,40]. This literature is extensive and covers resilience mechanisms [116], psychological capital and emotional intelligence as serial mediators of professional well-being [117], perfectionism and distress [118], emotional abuse and depression under moderated mediation [119], personality change following intervention [120], psychological need satisfaction as a predictor of exercise behavior [121], affective states as antecedents of performance-relevant experience [122], and the incremental validity of competing trait structures for explaining distress [123]. Organizational variants model psychological safety, embeddedness, and leadership effects on behavior [124–128], link psychological contracts to safety behavior in high-risk industries [129], and test through cross-cultural mediation designs whether such structures transfer across national contexts [130]. Nearly all of this cluster is E2: single-sample, cross-sectional, self-report. These models deliver explicitness, since every path is a stated hypothesis and the model is falsifiable in a way that a black-box predictor is not. They break on causal identification. Most are estimated on cross-sectional self-report, so the arrows encode assumed rather than demonstrated causal direction, a limitation this literature states less often than it should and one that should be applied when interpreting any effect size reported from it.

Hybrid structural–neural models and causal identification. Hybrid designs chain a structural model to a neural network, using the structural part to name the constructs and paths and the network to fit nonlinear relations among them [131] (E2). The first version of this review described this as a trade of explicitness for predictive power. Two reviewers objected, and they are right: the weakness of the structural family is not a lack of nonlinearity but a lack of identification, and adding a neural network to a cross-sectional path model changes nothing about endogeneity, reverse causation, or omitted confounding. It improves the fit of the same unidentified structure and, by making the functional form opaque, makes the assumption failures harder to see. We therefore restate what such hybrids can and cannot support. They can support a claim about prediction, that behavior can be forecast from the named antecedents more accurately than by a linear path

model, and that claim is tested by held-out prediction. They can support a causal claim only to the extent that the structural part is identified by design, and the identification comes from the design, not from the estimator. The designs that do this are the ones the causal-inference literature specifies [132,133]: a graph in which the assumed absence of unmeasured common causes is stated and defended; experimental or quasi-experimental variation in the antecedent, such as a randomized intervention on self-efficacy or a natural experiment in exposure; longitudinal designs that separate within-person from between-person components, since the conventional cross-lagged panel model does not by itself identify reciprocal effects [134]; and instruments or discontinuities where they exist. When a hybrid model rests on none of these, its paths should be read as conditional associations with a nonlinear fit, and the text and tables of this review label them accordingly. A hybrid that does rest on such a design can be more useful than either component alone, because the network can then estimate a nonlinear dose–response along a path whose direction the design has fixed; that, and not the addition of capacity, is the legitimate case for the combination.

Behavioral sequence and trajectory models. Where longitudinal behavioral traces exist, capability can be inferred from temporal structure directly. Temporal models of student behavior that are sensitive to anomalous patterns have been used for mental-health prediction [135], cognitive and non-cognitive factors jointly predict academic performance [136], and reinforcement-learning formulations have been applied to the cultivation of decision behavior in training contexts [137] (all E2). The gain is that capability is measured from what people do repeatedly rather than from what they report; the cost is that behavioral traces are typically collected in a single institutional context, so the estimated capability may be inseparable from the environment that elicited it, which is an instance of the generalization problem taken up in D3.

Table 6. D2—Individual behavioral capabilities: model families compared, with representative work and highest evidence tier.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
SEM/mediation models	Capability mediates between stated antecedents and outcomes	Cross-sectional survey, moderate n	Causal direction assumed, not identified	[116,117,124] (E2)
Multilevel/longitudinal SEM	Effects nest within persons and contexts	Repeated measures	Attrition; non-invariance; cross-lagged paths not identified without within-person separation	[128,130,134] (E2)
Hybrid structural–neural	Named paths with nonlinear fit; identification comes from design, not estimator	Survey plus design-based variation	Opaque fit of an unidentified structure	[131,133] (E2)
Behavioral sequence models	Capability is revealed by temporal behavioral structure	Dense behavioral logs	Capability confounded with eliciting environment	[135,136] (E2)
Reinforcement-learning formulations	Capability is a learned policy under feedback	Interaction traces or simulator	Cognitive implausibility of learning trajectory	[38,137] (E2)
Generative (LLM) agents	Capability can be instantiated and enacted	Persona specification; human comparison data	Believability mistaken for validity	[138–140] PP (E2; E1 for the 1052-agent result)

Generative agents as capability models. The newest entrant instantiates capabilities directly. LLM-driven agents with memory, planning, and reflection reproduce believable individual behavior over extended interaction [138], controlled studies show that such

agents can replicate classical human-subject findings [139], and interview-grounded agents built for 1052 real individuals reproduce those individuals' General Social Survey responses about 85 per cent as accurately as the individuals reproduce their own answers two weeks later [140]^{PP} (E1; the result is a preprint and has not, to our knowledge, been independently replicated). This inverts the dominant use of the rest of D2 and therefore inverts the validation burden, since the question is no longer whether parameters are identified but whether generated behavior matches human behavior distributionally, which requires held-out human data as the comparison standard [141]. Matching aggregate distributions is not sufficient, because persona-conditioned models can flatten within-group variation even while matching group means [142].

4.3. D3: Individual–Environment Interaction

In D3 the context c becomes a modeled variable in its own right rather than a nuisance term, and the object is the closed loop between situation and psychological response. Table 7 summarizes the existing modeling methods within this dimension.

Stimulus–organism–response formulations. The SOR schema, inherited from environmental psychology [41], remains the workhorse for representing how environmental cues produce internal states that in turn drive behavior. It has been applied to retention decisions under work–life demands [143], to consumer resilience under social-capital conditions [144], and to resident tourism-consumption behavior with neural-network estimation of the organism stage [145] (E2). Environmental-exposure studies extend the same logic to physical stressors by modeling how traffic noise at intersections produces annoyance and downstream psychological health effects [146], and the appraisal-based stress-and-coping tradition supplies the theoretical account of why identical exposures produce heterogeneous responses [28]. The formulation is interpretable and easily estimated, but it is typically static, capturing a single pass through the loop rather than the loop itself, and it inherits the identification problem of Section 4.2 when estimated cross-sectionally.

Domain-specific interaction models. Where the environment is well characterized, richer mechanistic models become possible. Driving research couples psychological state to control behavior through psychological-safety-space and style-coupled formulations [147,148], and hybrid pipelines that combine cognitive psychology with machine learning now assess driver situation awareness in conditionally automated vehicles, a setting in which mis-estimating the human's mental state has direct safety consequences [149] (E2, simulator studies). Technology-use research models the interaction of person, affect, cognition, and execution in the development of specific internet-use disorders and provides one of the few explicitly dynamic person–environment accounts in this class [150]. Human–robot interaction studies close the loop in real time by using detected affect to modulate the machine's behavior [107]. Reinforcement learning provides the general formal template for such closed loops, with state, action, environment, and reward [38], but psychologically grounded reward functions remain rare, which is why most deployed interaction models are still single-pass.

Decision under environmental uncertainty. A substantial applied literature models how environmental risk structure shapes psychologically realistic choice. Prospect-theoretic value functions supply the foundational departure from expected utility [33,34], and contemporary work embeds psychological risk attitudes in fuzzy and multi-source decision models for venture capital, product ranking, and insurance choice [151–154] (E2; evaluated by within-sample fit or by numerical examples). Large-group emergency decision-making under risk measurement extends this to collective settings and forms a natural bridge to D5 [155]. Foundational discussions of the relation between rational and psychological decision theories remain useful for situating these variants [156].

Table 7. D3—Individual–environment interaction: model families compared, with representative work and highest evidence tier.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
SOR path models	Environmental cue → internal state → behavior, single pass	Survey with cue manipulation	Static; does not model the loop; cross-sectional identification	[41,143,145] (E2)
Domain-specific mechanistic models	Environment is well characterized and low-dimensional	Instrumented domain data	Does not transfer beyond the modeled domain	[147,149,150] (E2)
Risk/decision models (prospect, fuzzy)	Choice reflects psychologically distorted value and probability	Choice data with known payoffs	Parameters absorb unmodeled context	[33,152,153] (E3 for the parametric core; E2 for applied variants)
Reinforcement learning	Adaptation is reward-driven policy improvement	Simulator or dense interaction	Impoverished environment yields brittle policy	[38,137] (E2)
LLM-in-the-loop ABM	Open-ended context can be supplied linguistically	Scenario specification; validation data	Inherits base-model cultural distribution	[17,18] (E2)

Where D3 breaks. Two failures recur. The first is environmental impoverishment: the modeled environment is usually a small set of hand-specified variables, so agents adapt to a caricature of the world and generalize poorly outside it. The second is cultural narrowness: because environments are sampled where data collection is convenient, person–environment couplings estimated in one cultural setting may not hold in another, which is the mechanism behind the generalization failures discussed in Section 7. LLM-in-the-loop agent simulation is the main current attempt to escape the first failure by supplying open-ended context [17], and it inherits the second.

4.4. D4: Individual Dynamic Change

D4 replaces the static latent z with an evolving z_t and models its trajectory, covering emotional evolution, learning and forgetting, development, and stress recovery. Two structural facts define the class. Conceptually, within-person dynamics cannot in general be recovered from between-person cross-sections, which is the ergodicity problem that Molenaar’s idiographic manifesto placed at the center of psychological methodology [62]. Practically, the requisite intensive longitudinal data come almost entirely from ecological momentary assessment and passive sensing [61,93], so measurement burden is highest here and methods are judged largely by their data efficiency. A third fact, raised by a reviewer, is that psychological dynamics are open rather than closed: the state is driven by an environment that is itself changing, so a law of motion estimated in one context is a conditional law, and the families below differ in whether they represent the environmental input explicitly (system dynamics and reinforcement-learning formulations do; growth and autoregressive models usually do not).

Latent-state and latent-growth models. Latent growth curve modeling estimates individual trajectories and their predictors and has been used to characterize the development of psychological flexibility [19] (E2). Hierarchical autoregressive extensions capture individual differences in dynamic characteristics rather than only in level and slope [157], which addresses a long-standing limitation of growth models, namely that two people with identical mean trajectories may differ entirely in volatility and in how quickly perturbations decay.

Dynamical-systems and early-warning formulations. Treating psychological state as a dynamical system yields predictions that static models cannot express. The most striking

is critical slowing down: as an individual approaches a transition such as depression onset, perturbations decay more slowly and autocorrelation and variance rise, which provides early-warning signals validated in intensive longitudinal data [60] (E3). The network-psychopathology program gives this a mechanistic reading, with disorders as alternative stable states of a symptom network [32,43], which makes D4 the class where formal theory and intensive longitudinal data connect most directly.

System-dynamics models of psychological stocks. The Forrester–Sterman framework represents a psychological process as stocks that accumulate and deplete through flows governed by feedback loops with delays [10]. At the individual level it has been used to formalize major depressive disorder as a set of interacting loops among stress, rumination, sleep, and activity, built from the verbal clinical literature and evaluated against reference behavior modes [11], to model the hypothalamic–pituitary–adrenal axis with an explicit stock of circulating cortisol [12], and to represent self-regulation, goal striving, and resource depletion in organizational research [13]. What SD delivers in this class is a transparent loop structure that can be inspected, a natural representation of delays and accumulation that autoregressive models lack, and a validation practice, reproduction of reference modes and structural tests, that is explicit even when data are sparse. What it forfeits is individual-level estimation: SD models in this literature are usually calibrated to typical or aggregate trajectories rather than fitted to each person’s time series, so they describe a representative person’s loop structure rather than any particular person’s, and they are evaluated at E2 by reproduction of qualitative behavior modes rather than by held-out prediction. The family is nevertheless the one that most directly bridges the verbal theories of Section 3.1 and the formal models of this class, because a causal-loop diagram is a disciplined way to turn a verbal theory into an *F3* specification (Section 4.7).

State-space and regulation models. Hidden Markov formulations remain the transparent baseline for discrete-state evolution [103,104]. A second line specifies regulation itself as a differential process of accumulation and dissipation, and its Bayesian multilevel implementation is notable for targeting the regime that constrains this entire class, since it returns relatively unbiased estimates from samples as small as about fifteen individuals or time series as short as about fifteen observations [20] (E2, parameter recovery in simulation). Because intensive longitudinal designs are expensive and attrition-prone, data efficiency is not a convenience in D4 but the binding constraint on what can be modeled at all, and neural approaches to psychological dynamics are evaluated on the same criterion, with simulation-based inference used to fit cognitive models whose likelihoods are intractable [158].

Theory-informed temporal networks. The most promising direction couples a psychological process theory to a temporal neural architecture. Psychology-informed neural networks apply memory-theoretic constraints to behavior prediction and improve both accuracy and interpretability relative to unconstrained sequence models [48] (E2, held-out prediction within the same learning platform), in direct analogy to physics-informed learning [51] and to the broader program of informed machine learning [52]. Applied work on resilience trajectories pursues the same data-efficiency goal with neurologically motivated architectures [159]^{PP} (E1).

Where D4 breaks. Longitudinal psychological data are noisy, irregularly sampled, and frequently missing not at random, and models that assume regular sampling degrade sharply under these conditions. Cross-sectional validation of a model intended to capture change, which is still common, provides essentially no evidence about temporal generalization, and we regard this as one of the most consequential evaluation gaps in the field. Table 8 summarizes the existing modeling methods within this dimension.

Table 8. D4—Individual dynamic change: model families compared, with representative work and highest evidence tier.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
Latent growth models	Change follows a parametric trajectory shared in form across persons	3+ waves	Misses volatility and perturbation decay	[19] (E2)
Hierarchical autoregressive models	Persons differ in dynamics, not only in level and slope	Intensive longitudinal	Demanding sampling; sensitive to irregular spacing	[62,157] (E2)
Hidden Markov/state-space	State evolves over a discrete or low-dimensional latent space	Repeated observations	State count chosen, not identified	[103,104] (E2)
Dynamical-systems/early-warning models	State has alternative attractors; transitions show critical slowing	Dense EMA time series	Signals require sampling faster than the dynamics	[32,60] (E3)
System dynamics (stock–flow)	State is a stock driven by flows, delays, and feedback loops	Reference behavior modes; sparse aggregate series	Calibrated to a representative person, not fitted per person	[11–13] (E2)
Bayesian reservoir/nonlinear dynamics	Nonlinear dynamics recoverable at modest series length	Tens of repeated observations	Interpretability of learned dynamics limited	[20,158] (E2)
Theory-informed temporal networks	A process theory constrains the learned temporal map	Behavioral traces plus theory	Constraint mis-specification degrades both goals	[48,52] (E2)

4.5. D5: Group Psychology

D5 addresses how individual psychology aggregates into collective phenomena, from the dyad upward. It is the field’s hardest micro–macro problem and the class in which simulation is the dominant use. Table 9 summarizes the existing modeling methods within this dimension.

Dyadic and small-group inference. The smallest collective is the pair, and a distinct literature estimates properties of pairs and small groups from interaction traces. Physiological synchrony and inter-brain coupling are estimated from simultaneously recorded signals in hyperscanning designs, whose validity as a method has been reviewed [160]; emergent leadership and dominance are inferred from nonverbal behavior in small-group meeting corpora with held-out evaluation [161] (E3); and shared affect in naturalistic conversation is annotated and modeled at the level of the dyad [162]. These studies are inverse in use, since the collective property is estimated from the interacting units’ data, and they are the clearest counterexamples to any claim that group-level modeling is simulation-only. They are also the level at which multilevel data are most often available, which matters for claim C3 (Section 5).

Interdependence and multilevel models. Statistical approaches extend individual models to account for non-independence within groups. Actor–partner interdependence formulations model person–group effects explicitly [59], and applied studies model group cohesion as an emergent property of training structure [163] or trace team performance to dyadic exchange quality and shared psychological empowerment [164] (E2). These preserve the interpretability of the structural models of D2 while respecting nesting, but they characterize aggregate association rather than generative mechanism.

Mechanistic models of collective dynamics. A separate tradition, consolidated in the statistical physics of social dynamics [56], specifies interaction rules and derives macro behavior from them. Social-force models of pedestrian dynamics remain the reference formulation for physical crowd behavior [42], validated against escape-panic phenom-

ena [165] and refined by experiment-grounded heuristics [166] (E3 for the physical formulation, which has been calibrated against observed crowds); contemporary evacuation research extends them with explicit psychological terms such as impatience contagion and panic [167–171] (E2; the psychological terms are evaluated by simulation). Opinion dynamics runs on the parallel track of bounded-confidence and influence models [172], now coupled into multi-attribute group decision-making [173]. Epidemic-style formulations model the spread of emotion, rumor, and opinion, including multi-platform emotional contagion [174], stochastic rumor persistence and extinction [175], and coupled information-propagation dynamics [176] (E2); the contagion premise itself has direct empirical support of unequal strength: for behavioral spread it comes from a randomized field experiment in an online network [57] (E4), whereas for emotion it comes from observational measurement of contagion in social-media streams [177] (E2), so the emotional premise of these models currently rests on correlational rather than design-identified evidence. Related formal work analyzes nonlocal cooperative behavior and collective decision-making [178], and consensus and quantum group-decision models incorporate psychological preference into aggregation [155,179,180]. System-dynamics models sit in the same tradition at the aggregate level: stock–flow models of population mental health and of the diffusion of behavior represent the collective as accumulating stocks linked by feedback, and the comparison of SD with agent-based formulations of the same process shows when heterogeneity and network structure change the aggregate outcome [14,15]. This family delivers micro-to-macro derivation, in which the emergent pattern is explained by the rules rather than merely fitted, at the cost of drastically simplified agents.

Table 9. D5—Group psychology, from the dyad upward: model families compared, with representative work and highest evidence tier.

Model Family	Core Assumption	Data Need	Characteristic Failure	Repr. Work (Tier)
Dyadic/small-group inference	Interaction-borne property is estimable from the interacting units' signals	Simultaneous multi-person recordings	Small samples; laboratory settings; synchrony metrics not standardized	[160–162] (E3)
Interdependence/multilevel models	Group membership induces non-independence	Nested survey data	Association, not generative mechanism	[59,164] (E2)
Social-force/physical crowd models	Collective motion follows interaction forces	Trajectory or flow data	Psychological terms added post hoc	[42,166,167] (E3 physical core; E2 psychological terms)
Opinion-dynamics models	Attitudes update through local influence rules	Attitude time series; network structure	Influence rules rarely identified from data	[172,173] (E2)
Compartmental (SIR/SEIR-type) contagion	Emotion or opinion spreads like an infectious process	Aggregate time series	Many rule sets fit the same macro curve	[174,175] (E2; contagion premise E4 for behavioral spread via [57], E2 for emotion via [177])
System dynamics (stock–flow)	Collective is a set of aggregate stocks linked by feedback	Aggregate series; reference modes	Heterogeneity and network structure averaged away	[14,15] (E2)
Rule-based ABM	Macro pattern is derivable from micro rules	Calibration targets	Agents drastically simplified	[58,181] (E2)
Generative (LLM) social simulation	Rich individual behavior improves macro realism	Scenario design; multilevel validation data	Validation weak; base-model bias amplified	[17,138,140] PP (E2; E1 for population-level claims)

Agent-based and generative social simulation. Agent-based modeling relaxes the analytic tractability constraint and permits heterogeneous agents [58,181], with a dedicated literature linking cognitive psychology to collective behavior simulation [67]. Psychological models of collective risk perception apply the same logic to how individual judgments aggregate into shared risk estimates [182]. The current frontier replaces rule-based agents with LLM-driven ones, producing emergent conventions and information cascades with far richer individual behavior [17,18,138,183], a development social scientists have greeted as potentially transformative [184,185]. Whether such populations reproduce real collective structure is an open empirical question. Interview-grounded agents approximate individual survey responses [140]^{PP} and LLMs can approximate regional psychological structure [186]^{PP} (both E1, but conditioning can equally distort subpopulation responses [141], flatten within-group heterogeneity [142], and synthetic populations may inherit and amplify the base model's demographic skew [114]).

Where D5 breaks. The binding problem is validation. Macro-level plausibility is weak evidence, because many different micro-rule sets generate qualitatively similar macro patterns, and distinguishing among them requires calibration against real collective data at more than one level simultaneously, which the datasets of Section 5 rarely support. This is claim C3 in its D5 form. Until multilevel validation data exist, claims that a simulated society reproduces a real one should be treated as hypotheses rather than as results. The stakes extend beyond methodology, since collective-behavior modeling increasingly informs platform design and public policy and has been described as a crisis discipline in which poorly validated models can do real harm [187].

4.6. Cross-Class Synthesis

Reading the five classes together yields three observations that no single class makes visible, and they correspond to claims C1 to C3.

First, the trade-off between transparency and capacity is not monotone across classes. In D1 and D4, where constructs are well theorized and measurement traditions are strong, transparent latent-variable, process, and stock–flow models remain competitive with deep alternatives, and theory-informed hybrids [48] sit on the efficient frontier, even as foundation models of cognition set a new predictive ceiling whose explanatory status is contested [53,54]. In D3 and D5, where the environment and the interaction structure are themselves poorly specified, high-capacity generative methods are the default, not because they are better understood but because no adequate structural alternative exists. The first version of this review inferred that the balance should shift toward structural models as interaction theory improves. A reviewer pointed out that this neglects a second possibility: the targets of D3 and D5 may be context-dependent, open, and multiply realizable in ways that make closed theorization of the D1 kind unattainable in principle, so that high-capacity methods are the right default rather than a stopgap. We now treat the two readings as competing hypotheses. The first predicts that, as interaction theories are formalized, theory-informed hybrids in D3 and D5 will show the same capacity-matched gains that they show in D1 and D4; the second predicts that they will not, because the constraints available will be too weak or too local. The evidence to decide between them does not yet exist, and the readiness rubric of Section 3.1 is the instrument for collecting it.

Second, validation practice follows the dominant use of a model rather than the most severe test the data permit. The estimation classes inherited a mature validation apparatus from psychometrics, comprising construct validity, reliability, and measurement invariance [44,50], and forward-form models used in those classes are validated by parameter recovery and held-out prediction. The simulation use, dominant in D5 and growing in D3, has no equivalent settled standard, and the criteria in current use, namely face plausibility and replication of a

handful of classic findings, are weaker than what the same researchers would require of an estimated model. The dyadic and small-group literature shows that this is not a property of the collective level as such, since collective properties estimated from interaction traces are validated with held-out data like any other estimand. Importing psychometric validation logic into generative simulation, in the form of held-out human data as the comparison standard, tests of within-group heterogeneity, and pre-registered predictions of collective outcomes, is on our reading the single most valuable methodological transfer available to the field, and it is the practical content of claim C2.

Third, the micro–macro link is a data problem before it is a modeling problem. Every class contains capable methods, but the quantities that would decide between aggregation operators, or separate individual susceptibility from structural exposure, are cross-level parameters, and identifying them requires the same people to be observed as individuals, in their interactions, and in the collective outcome. Section 5 lists the resources that observe two of the three and states what the third would require. This is claim C3, and it is a resource gap rather than an algorithmic one.

4.7. Boundary Conditions and Counterexamples for C1–C4

The four claims were defended above by supporting cases. This subsection, added at the reviewers' request, states for each claim the cases that count against it, the domain over which we still hold it, and the observation that would lead us to withdraw it.

C1: Readiness determines feasibility. Counterexamples of three kinds exist. First, theory-informed modeling with verbal (F1) theories has been successful when the theory is used at the structure level rather than the function level: the entire structural-equation literature of D2 constrains a model with self-determination or social-cognitive theory, and system-dynamics models of depression and of goal striving turn verbal clinical and organizational theories into executable loop structures [11,13]. Second, LLM persona conditioning uses verbal trait descriptions as constraints at the prompt level and produces measurable shifts in generated behavior [141]. Third, in the case of risky choice, readiness was created by data rather than presupposed: a machine-learned value function outperformed every parametric theory and was then read back as theory [85]. These cases fix the boundary. C1 is a claim about theory-informed *learning*, in which a theory is embedded as a constraint inside a learned function class at the representation layer; it is not a claim about theory-informed modeling in general, which includes structure-level, prompt-level, and hand-built formalizations. Within that boundary, the cases in which a verbal theory has constrained a learned function class have first required someone to write a functional form the theory did not state, which is the step that a causal-loop diagram makes explicit and that the rubric scores as moving from F1 to F3. The claim would be falsified by capacity-matched theory-informed hybrids in F1 domains that show the same out-of-distribution gains as those in F3 domains without such an intermediate formalization step. Because the readiness scores are rubric-based judgments rather than instrument measurements, C1 should be read as a well-specified conjecture with a stated test, not as a result.

C2: The shift from estimation to simulation. The counterexamples raised by a reviewer are real, and they changed the claim. Forward-form generative models (drift-diffusion, reinforcement learning, ACT-R) dominate parts of D1; inverse estimation of cohesion, collective affect, leadership emergence, and network parameters from interaction traces is common at the dyadic, small-group, and group levels (Section 4.5); and Table 3 records both. The revised claim is therefore not about model form but about the dominant use and its validation target, and its domain is the generative-simulation literature of D5 and D3 (rule-based and LLM-driven agent models, mechanistic crowd and contagion models with psychological terms) rather than the collective level as a whole. Within

that domain, the claim would be falsified by a body of simulation studies that estimate their micro-rules from the interacting units' own data and validate them against held-out collective outcomes; the dyadic hyperscanning and small-group literature shows that this is possible, which is why we regard the gap as a practice gap and not a level property.

C3: Data, not model capacity, is the binding constraint. The first version of this review stated that “almost no resource” observes the same people as individuals and as members of a collective. That was too strong. Resources exist that record individual states together with interaction structure over time: longitudinal sensing studies of student cohorts [188,189], wearable proximity networks [190,191], sociometric-badge studies of teams [192], annotated meeting corpora [161,193], dyadic hyperscanning and conversation datasets [160,162], and dyadic interaction corpora such as IEMOCAP [194]. The revised claim is that these resources are few, short (weeks to months, rarely years), small (tens to hundreds of people), concentrated in student and laboratory populations, and, most importantly, rarely record the collective outcome that a simulation targets, such as a decision the group actually made, a norm that shifted, or a cascade that occurred, at the same time as the individual states and the interaction structure. The quantities that remain unidentified are stated in Section 5. C3 would be falsified by the existence of a resource with all three components at matching timescales on which competing aggregation operators had been discriminated; we are not aware of one, and we would welcome correction. C3 does not claim that model capacity is irrelevant, only that additional capacity cannot substitute for the missing observations.

C4: The evidence base is heterogeneous in quantity and quality. The claim has two boundary conditions. First, the gradient is a property of the evidence for a finding, not of a model family: social-force crowd models have tier E3 evidence for their physical core and tier E2 evidence for their psychological terms, and the same distinction applies to any hybrid. Second, the count of one randomized trial of a generative intervention [195] is a statement about the included set at the time of the search; further trials are likely, and the claim is that readiness should be indexed to the tier of the available evidence, whatever its current level, rather than that the level is fixed. C4 would be undermined, in the useful sense, by a field in which external validation and design-identified evidence became routine, and the evidence-tier coding of Section 2 is offered as a way of tracking that.

5. Datasets

Models become actionable only through data. This section surveys widely used resources and their modeling roles and then states their systemic limitations, the fourth of which is the evidence for claim C3. Because dataset availability, licensing, and naming change quickly, we describe categories and representative resources rather than asserting fixed specifications, and we recommend verifying provenance and licensing before use. Tables 10 and 11 organize representative datasets by the primary construct, the unit of analysis and interaction structure, and the observable, following the reviewers' request that constructs and observables be separated and that each dataset's unit and interaction structure be stated. Three classifications in the first version were wrong and are corrected here: DREAMER is an emotion-recognition dataset with EEG and ECG observables and is now listed under emotion; AffectNet is a facial-expression dataset annotated for categorical emotion and valence–arousal and is now listed under emotion rather than mental health; and IEMOCAP is a dyadic interaction corpus and is now listed with its interaction structure stated. A category of dyadic, small-group, team, and longitudinal-network resources has been added, because the first version's claim C3 was made without having searched for it.

Table 10. Representative datasets for psychological and behavioral modeling, grouped by primary construct, with unit of analysis, interaction structure, and observable stated for each.

Representative Resources	Unit; Interaction Structure	Observable	Primary Modeling Role
Individual: cognition, decision, and reasoning			
Psych-101 [53], H-ARC [196] PP	Person; none	Trial-level choices, response times	Decision, memory, and reasoning; cognitive process and foundation models
Individual: emotion and affect			
DEAP [197], SEED [198], FACED [199], DREAMER [200]	Person; none (stimulus-elicited)	EEG, ECG, peripheral physiology	Physiological emotion recognition (categorical and valence–arousal)
AffectNet [201]	Person; none (in the wild)	Facial images	Categorical and dimensional expression recognition
RAVDESS [202], CMU-MOSEI [203]	Person; none	Speech, audio-visual, text	Speech and multimodal affect and sentiment
IEMOCAP [194]	Person within dyad; scripted and improvised dyadic interaction	Audio-visual, motion capture	Emotion in conversation; interlocutor context
Individual: personality			
PANDORA [204], MBTI-9k [205], Twitter personality corpora [206,207]	Person; none	Social-media text	Trait–language association; Big Five and MBTI prediction
Individual: mental health and stress			
DAIC-WOZ [208]	Person; interview with a virtual agent	Speech, language, facial features	Distress and depression screening; consent and de-identification required
WESAD [209]	Person; none (protocol-elicited)	Wearable physiology	Stress and affect detection under protocol

Table 11. Representative datasets: dyadic, small-group, team, and population-level resources, with unit of analysis, interaction structure, and observable stated for each.

Representative Resources	Unit; Interaction Structure	Observable	Primary Modeling Role
Dyad, small group, and team: interaction, synchrony, and longitudinal networks			
K-EmoCon [162]	Dyad; naturalistic debate	Audio-visual, physiology, self and partner ratings	Shared affect; dyadic emotion recognition
AMI [193], ELEA [161]	Small group (3–4); meetings	Audio-visual, transcripts, nonverbal features	Emergent leadership, dominance, group decision process
Hyperscanning datasets (reviewed in [160])	Dyad; joint task	Simultaneous EEG/fNIRS	Inter-brain and physiological synchrony
Sociometric-badge team studies [192]	Team; face-to-face networks over weeks	Proximity, speech activity, motion	Team interaction structure and performance
StudentLife [188], Copenhagen Networks Study [189], Reality Mining [191], SocioPatterns [190]	Person within cohort; longitudinal proximity and communication networks (weeks to months)	Smartphone sensing, Bluetooth/RFID proximity, calls, EMA	Individual states with interaction structure over time; the closest existing approximation to the resource claim C3 calls for
Population: values, well-being, and cross-cultural			
World Values Survey [210], ESS [211], CGSS [212], European Values Study [213]	Population; repeated cross-sections	Survey	Social attitudes and value orientation across regions and waves
World Happiness Report [214], HRAF [215]	Population; none	Survey, archival	Well-being; cross-cultural comparison
Social-media sentiment corpora [91]	Population; online networks	Text, behavioral traces	Public-emotion monitoring and early warning

Resources are indicative rather than exhaustive; availability, licensing terms, and documented biases should be confirmed per project before use.

5.1. Individual, Interaction, and Population Resources

Five categories recur at the individual level. Cognitive and behavioral resources support process modeling, and their scale has changed qualitatively: Psych-101 aggregates trial-by-trial data from 160 experiments, 60,092 participants, and more than ten million choices into a single natural-language format [53], while carefully normed small resources such as H-ARC provide human baselines for reasoning benchmarks [196]^{PP}. Emotion corpora span physiological signals, including DEAP, SEED, the finer-grained FACED, and the consumer-grade DREAMER [197–200], facial expression in the wild [201], and audio-visual and conversational material such as RAVDESS, IEMOCAP, and CMU-MOSEI [194,202,203]. Personality corpora annotate social-media text for Big Five or MBTI labels, from early Twitter collections [206,207] through the Reddit-based MBTI-9k [205] to the more comprehensively annotated PANDORA [204]. Mental-health corpora such as the DAIC-WOZ distress interviews require particular care over consent and de-identification [208], and wearable resources capture stress under protocol [209].

At the interaction level, the resources added in this revision fall into three kinds. Dyadic corpora record two people in conversation or joint task with simultaneous physiology or brain signals, which supports estimation of shared affect and synchrony [160,162]. Small-group and team corpora record meetings and work teams with audio-visual, non-verbal, and proximity observables, which supports estimation of emergent leadership, dominance, and team interaction structure [161,192,193]. Longitudinal cohort studies record individual states, through EMA and smartphone sensing, together with communication and proximity networks over weeks to months [188–191]; these come closest to the multilevel resource that claim C3 calls for, and their limitations are stated below. At the population level, value and social-attitude surveys [210–213], well-being resources [214], social-media sentiment corpora [91], and cross-cultural archives [215] support the analysis of collective psychology at the extension level of Section 3.3.

5.2. A Structured View and Its Limitations

Datasets bridge abstract constructs and observed behavior and thereby enable the shift from theoretical hypothesis to prediction. Five systemic limitations constrain reliability, and they recur as the bottlenecks of Section 7: (i) *static self-report bias*, where cross-sectional scales miss the high-frequency dynamics that only ecological momentary assessment captures [61]; (ii) *fusion difficulty*, where heterogeneous modalities differ in resolution and noise [105]; (iii) *imbalance and cultural bias* arising from sampling and annotation, which is the WEIRD problem in dataset form [216] and which skews both performance and cross-cultural transfer; (iv) the *micro–macro divide*, in which most datasets isolate either individual cognition or collective behavior; and (v) *privacy and ethics*, where the tension between anonymization and the resolution that models require remains unresolved, since even heavily incomplete records support re-identification [217].

The fourth limitation is claim C3 in its revised form, and it can now be stated precisely. A multilevel resource adequate for validating a cross-scale model needs three components on the same people: (a) individual states sampled faster than they change, which for affect and stress means EMA or passive sensing at hours-to-days resolution [60]; (b) the interaction structure at the resolution of the interaction, which for a team means who spoke to whom in each meeting and for an online cohort means the communication network at daily resolution; and (c) the collective outcome that the model targets, such as a group decision, a consensus reached, a norm adopted, or a cascade observed, time-stamped so that it can be aligned with (a) and (b). The longitudinal cohort studies in Table 11 supply (a) and (b) for tens to hundreds of people over weeks to months; the meeting corpora supply (b) and, for single sessions, (c); the hyperscanning and dyadic corpora supply (a) and (b)

for pairs in the laboratory. We are not aware of a resource that supplies all three at matching timescales for the same people, and that is the sense in which the claim is now made. What such a resource would make identifiable is stated in Section 3.3: the aggregation operator, the split between individual susceptibility and structural exposure, the cross-level effect of an individual-level intervention on the collective outcome, and the timescale of the downward path. Without it, generative social simulation cannot be validated at more than one level simultaneously, and it is for that reason, rather than as a general preference for data over methods, that we regard multilevel, consent-governed longitudinal collection with all three components as the highest-value investment available to the field. Any such collection must be designed with the privacy limitation (v) in view, because it is precisely the linking of individual and relational records that makes re-identification easy [217].

6. Multi-Domain Applications

This section examines deployment across four civilian domains, namely mental health, education, decision support, and emergency management, and treats military use as a bounded case with distinct ethical constraints (Section 6.5). A reviewer asked why these domains and not others. The four civilian domains were selected on three stated criteria: each is a setting in which a psychological model is used to make or support a decision about real people, rather than only to describe them; each is represented in the included set by a cluster of applied studies large enough for a statement about evidence quality to rest on the cluster rather than on a single paper; and, taken together, they span the five classes and the full range of evidence tiers. Other application areas are present in the review but are treated as sources of modeling studies rather than as deployment domains: human–computer and human–robot interaction and driving appear in D3 (Section 4.3), sports and organizational behavior in D2, marketing and consumer choice in D3, and public health in the system-dynamics and contagion families of D4 and D5. The domains also make claim C4 concrete, because the strongest evidence available differs sharply across them, and Table 12 now states the highest evidence tier reached in each domain alongside the modeling classes, dominant model families, and main barrier to deployment.

Table 12. Application domains: modeling classes, dominant methods, strongest available evidence with its tier, and the main barrier to deployment.

Domain	Classes	Dominant Model Families	Strongest Available Evidence (Tier)	Main Barrier to Deployment
Mental health	D1, D4	Supervised classifiers; LLM assessors; generative interventions	One randomized controlled trial of a generative chatbot against a waitlist control [195] (E4); external validation of screening models remains rare	External validation absent for most classifiers; consent and privacy
Education	D2, D4	SEM and mediation models; theory-informed sequence models	Theory-constrained prediction with held-out gains on one learning platform [48] (E2); mediation structures are cross-sectional (E2)	Single-institution, cross-sectional designs; effect sizes do not transfer
Decision support	D3, D5	Prospect-theoretic and fuzzy models; consensus and opinion models; ABM	Within-sample fit, numerical examples, and simulation studies [181,182] (E2); the parametric core of prospect theory is E3	Parameters absorb unmodeled context; no out-of-sample testing in the applied cluster
Emergency management	D5	Social-force and cellular-automata models; compartmental contagion models	Physical formulations calibrated against observed crowd behavior [165,166] (E3); psychological extensions are simulation-only (E2)	Psychological terms added on top of physical ones and rarely identified separately
Military (bounded)	D1, D2	Assessment instruments; simulation of stress effects	Small-sample instrument validation [218,219] (E2)	Consent, dual use, accountability, and legal review

6.1. Mental Health

Mental health has the deepest evidence base, consolidated in scoping and clinical reviews of machine learning across screening, diagnosis, and treatment-outcome prediction [64–66]. For identification and diagnosis, explainable models predict depression, anxiety, and stress from task data [7], domain-tuned encoders flag patients who need therapy [220], hybrid deep models classify disorder types with high but dataset-specific accuracy [221,222], and machine learning models profile resilience and post-viral distress in specific populations [99,223] (all E2). The LLM layer is developing on two tracks. On the assessment track, professional item banks and competency benchmarks quantify the counseling ability and psychological knowledge of general and Chinese-language models [224–226] (E2; the counseling-ability benchmark is a preprint^{PP}, E1), dedicated psychological LLMs outperform general models in Chinese counseling [227,228] (E2; the Psy-LLM result is a preprint^{PP}), and instruction-tuned models produce interpretable mental-health analyses of social-media text [229,230] (E2). On the intervention track, proposals for responsible development [231] have been followed by the first randomized controlled trial of a generative-AI therapy chatbot, in which 210 adults with depression, anxiety, or high risk for a feeding or eating disorder were randomized to four weeks of chatbot access or to a waitlist control, with significant symptom reduction in the treatment arm [195] (E4). Further applied work monitors risk in specific populations with embedded-feature and attention-based models [232,233], gives early warning of distress through machine learning nomograms [234], and estimates resilience or detects negative affect with data-efficient and interpretable multimodal frameworks [3,159] (E2; the resilience model is a preprint^{PP}). The evidence gradient should be stated plainly. Single-dataset classification accuracy is abundant, external validation is scarce, and randomized evidence of benefit exists so far for one generative intervention against a waitlist rather than against an active comparator, so claims of clinical readiness should be indexed to that gradient rather than to benchmark scores.

6.2. Education

Theory-informed networks such as PsyINN improve memory-behavior prediction while enhancing interpretability, and they do so by constraining a neural model with classical memory equations rather than by adding parameters [48] (E2, held-out within one platform). Moderated-mediation analyses show how engagement and self-efficacy relate to flexibility and resilience through mediators such as language anxiety and emotion regulation [235,236], and cognitive and non-cognitive factors jointly predict academic performance [136] (E2, cross-sectional; the mediation paths are associations in the sense of Section 4.2). For evaluation and technology adoption, structural models validate career-readiness constructs [237], hybrid SEM and neural pipelines link psychological needs to the sustainability of immersive learning environments [131] (E2; the hybrid adds nonlinear fit, not identification), and case studies examine the conditions under which interdisciplinary learning succeeds [238]. For joint modeling of mental health and academic outcomes, explainable ensembles support stratified intervention [239] and latent-growth designs characterize trajectories of psychological flexibility [19] (E2). Most of this literature is single-institution and cross-sectional, so what transfers is the modeling pattern, namely theory-constrained prediction combined with an explicit mediation structure, rather than the reported effect sizes.

6.3. Decision Support

Two reviewers noted that the first version of this section did not connect its examples to the evidence gradient. Every study cited in this subsection is tier E2: it is evaluated by

within-sample fit, by numerical examples, or by simulation, and none reports out-of-sample testing on independent decisions or a real-world outcome. We state this once here and read the paragraph accordingly. Agent-based modeling generates collective phenomena from individual interaction and supports theory building [181], and psychological models of collective risk perception connect individual judgment to aggregate risk estimates [182]. Quantum group-decision, opinion-dynamics-aware, and cognitive-state consensus models are designed for high-uncertainty settings [173,179,180] and are demonstrated on constructed examples. In economic decisions, prospect-theoretic fuzzy optimization improves perceived-value allocation [154], hybrid choice models mine multi-factor purchase intentions [153], and deep models analyze consumer psychology in tourism markets [145] in single samples. In driving, psychological-safety-space and coupled psychology-and-style models improve lane-change and control decisions [147,148], situation-awareness recognition supports safe handover in conditional automation [149], and multi-source ranking models capture loss and regret aversion in consumer choice [152] in simulator or offline settings. Simulation and within-sample results of this kind show that a psychologically realistic model can be formulated and can fit the data it was built on; they do not show that it improves real decisions, and no study in this cluster claims to have tested that. This is why decision support appears in Table 12 with the weakest evidence among the civilian domains, and it is the domain in which the gap between formulation and demonstrated effectiveness is widest.

6.4. Emergency Management

The evidence in this domain has two layers, and the reviewers were right that the first version did not keep them apart. The underlying physical formulations of crowd motion are among the few models in this review calibrated against observed crowd behavior [165,166] (E3). Built on that core, impatience-contagion, mixed-logit, and extended social-force models quantify how psychological factors shape route choice and flow [167–169], and cellular-automata simulations compare updating strategies [170] (E2: the psychological terms are evaluated by the plausibility and sensitivity of simulated outcomes, not by calibration against evacuations in which the psychological variables were measured). For public opinion and epidemics, stochastic SEIR models give conditions for rumor persistence or extinction [175], and information-propagation models couple social psychology with differential-equation dynamics and deep learning [176] (E2, analytical and simulated). For subway incidents, latent-class models classify evacuees into psychological profiles and improve route-choice prediction [171] (E2, stated-preference survey). Two contributions complete the picture at the organizational level: psychological contracts link safety climate to safety behavior on high-risk sites [129] (E2, cross-sectional), and multimodal physiological and psychological assessment has been applied to injury risk in high-intensity emergency training [240] (E2). The characteristic weakness here is identification rather than fit, because psychological terms are typically added on top of a physical model whose parameters already reproduce the aggregate flow, so the aggregate evidence for the physical core cannot be read as evidence for the psychological extension. Simulation results in this domain should be read as demonstrations that a psychological mechanism is consistent with known crowd physics, not as demonstrations that the mechanism operates in real evacuations.

6.5. Military Use: A Bounded, Ethically Constrained Case

Military applications illustrate both the reach and the risks of psychological modeling and are treated here as a delimited case rather than as a general selling point. Reported work includes models of soldier psychological safety [218], measurement instruments for

military support and its psychological correlates [219], tiered psychological first aid for combat stress [241], and simulation of the effect of fear on action [242] (E2 at most). These uses raise concerns about consent, dual use, accountability, and the law of armed conflict that civilian domains raise less sharply. We include them to make the risk surface explicit and to motivate the governance discussion of Section 7.2, not to advocate deployment. Any real-world use in this domain requires independent ethical and legal review.

7. Open Challenges and a Research Agenda

The preceding sections expose five recurring bottlenecks. Table 13 maps them to modeling classes and to concrete research needs. Section 7.2 then addresses the five in the same order. For each it separates a *research-practice recommendation*, which is a statement about how work should be done and which no experiment can refute, from a *testable prediction*, which names a hypothesis, a measurable outcome, a decision criterion, and the result that would reject it. The first version of this review called all of these “falsifiable tests”; the reviewers correctly noted that several were governance or practice proposals, and the two kinds are now labeled as such.

Table 13. Research gaps and needs across the five modeling classes, with representative evidence.

Class	Main Gaps	Research Needs	Evidence
D1 States	Opaque models; weak theory integration; prediction without explanation	Theory-informed hybrids; explainable and auditable assessment	[8,48,53]
D2 Capabilities	Static cross-sectional designs; unidentified causal paths; unreliable individual-level measurement; cultural bias	Lifelong capability tracking; design-based identification; generative measurement models; bias-controlled assessment	[49,133,216]
D3 Interaction	Weak open-world generalization; over-simplified feedback	High-fidelity feedback loops; multi-agent co-adaptation	[17,150]
D4 Dynamics	Poor multiscale fusion; noise sensitivity; snapshots instead of processes; environment treated as closed	Robust multimodal alignment; noise-resilient tracking; early-warning monitoring; explicit environmental input as in stock–flow models	[11,60,62]
D5 Group	Neglects individual cognitive diversity; weak validation; dyadic level under-integrated; no human–AI co-evolution	Cognitively rich emergence; multilevel validation data; operator comparison; trustworthy human–AI decision loops	[15,140,142]

7.1. Challenges

7.1.1. Interpretability

Many high-performing deep and large models are opaque, so their decision logic cannot be verified or mapped onto psychological constructs, which undermines scientific credibility and safe use in high-stakes settings such as clinical assessment and personnel selection [7,8]. Post hoc attribution methods such as SHAP mitigate the problem without resolving it [243], and for high-stakes decisions the stronger position is that inherently interpretable models should be preferred outright [244]. The deeper issue is the divide between prediction and explanation, since a model can predict well while explaining nothing [54], and models validated by statistical fit alone produce predictions without theoretical support. Interpretability is not purely a cost. Analysis of model behavior can generate hypotheses about human cognition, as when machine-learned value functions recovered choice structure that classical parametric theories had not specified [85] and when cognitive-psychology batteries applied to language models produced testable characterizations of their decision profiles [111]. The exchange runs in both directions, but unevenly: theory-informed modeling, in which a theory constrains a learned model, is now an established practice, whereas prediction-informed theory revision, in which a learned

model's behavior is used to revise the theory, remains rare outside a few cases such as the value-function result. Widening the second direction is the first item of the agenda.

7.1.2. Privacy and Ethics

Collecting and integrating high-resolution psychological data creates serious privacy risk. The founding demonstrations that ordinary digital footprints predict sensitive traits [87] and support targeted persuasion [245] define the threat model, and current de-identification does not remove it, since incomplete datasets still permit successful re-identification with generative models [21,217]. The multilevel resources called for by claim C3 sharpen the risk, because linking individual states to relational records is exactly what makes a record identifying. Blurred responsibility among developers, users, and overseers raises the risk that model outputs displace expert judgment, and abstract ethical principles alone have proved insufficient to govern practice [246].

7.1.3. Cultural Bias and Generalization

Models trained on culturally narrow data generalize poorly. The problem predates machine learning, since behavioral science itself over-samples WEIRD populations [216] and over-relies on English [247], but modern pipelines industrialize it. Default opinions expressed by LLMs reflect identifiable demographic subgroups [114], psychometric adaptation across cultures remains unreliable [22], and decision models that ignore cross-cultural differences risk systematic error. The same narrowness limits transfer between settings within one culture, because person–environment couplings estimated in one context are rarely re-estimated in another (Section 4.3).

7.1.4. Temporal Dynamics

Psychological attributes are dynamic, yet most paradigms classify or regress static cross-sections and omit learning, forgetting, and adaptation, despite the formal argument that between-person snapshots cannot substitute for within-person processes [19,62]. Such models degrade under behavioral or environmental change and miss the clinically actionable transition signals that early-warning research has validated [60], and the scarcity of tools for lifelong tracking means that many assistive systems offer only short-term reference [159]^{PP}. A reviewer added a point that the first version missed: psychological dynamics are open. The state is driven by an environment that is itself changing, and the problem is most acute in exactly the classes where the dynamical approach is least used, such as the cross-sectional structural models of D2 and D3, which cannot represent an environmental input at all. Stock–flow and reinforcement-learning formulations represent the input explicitly; growth and autoregressive models usually do not, and a law of motion estimated without it is a conditional law whose domain of validity is the context in which it was estimated.

7.1.5. Human–AI Collaboration Risk

Emerging systems tend toward one-directional designs in which model output replaces human judgment, which neglects complementary strengths and ethical checks; formal models of repeated trust interaction show that reliance on an automated partner is itself a psychological process that can be modeled and therefore misdesigned [248]. A specific and underexamined risk is the use of LLMs as simulated participants. The prospect is attractive to social science [184,185], and conditioning can make model responses mirror human subpopulations, but it can equally distort them [139,141], misportraying and flattening the identity groups being simulated [142] and failing to reproduce even human response biases [115]. LLM-generated data should therefore be validated against human data rather than substituted for it [183].

7.2. A Research Agenda: Recommendations and Testable Predictions

The five directions below correspond one to one with the five challenges above. Each states a mechanism and a target; each then gives a research-practice recommendation and, where one can be stated with a measurable outcome and a rejection criterion, a testable prediction. Two of the five (privacy and human–AI collaboration) are governance matters for which no scientific hypothesis is at stake, and they carry recommendations only.

7.2.1. Theory-Informed and Explainable Modeling

Embedding psychological structure in learned models, as knowledge-guided networks such as PsyINN do, translates theory into constraints or symbolic modules by analogy with physics-informed learning and with the broader informed-machine-learning program [48,51,52]. The readiness gradient of claim C1 indicates where to start, namely with constructs that have *F3* theory, such as decision, memory, symptom dynamics, and stock–flow formulations of stress, before verbal-theory domains. *Recommendation*. Report every theory-informed model against an unconstrained baseline of matched capacity, on a held-out distribution that differs from the training distribution in a stated way, and report a construct-validity metric alongside accuracy. *Testable prediction*. Hypothesis: on a shared benchmark such as Psych-101, a model constrained by an *F3* theory will show higher out-of-distribution accuracy and higher construct validity than a capacity-matched unconstrained model, and the advantage will be absent for constraints derived from *F1* theories without an intermediate formalization step. Outcome: the difference in held-out accuracy and in the correlation between recovered parameters and independent measures of the construct. Criterion: a pre-registered minimum effect with a confidence interval excluding zero. Rejection: if the advantage disappears once capacity is matched, or if *F1*-derived constraints match *F3*-derived ones, the constraint is decorative and C1 is wrong in the case tested [53,55].

7.2.2. Risk Prevention and Accountability

Systematic data and algorithm governance requires complete authorization chains, a psychological-data responsibility framework with transparency standards that goes beyond principle statements [21,246], and privacy-preserving training such as federated optimization for sensitive domains [249]. *Recommendation*. Re-identification risk should be assessed as a standard step before any release of psychological data or of a model trained on it, with risk measured on the actual release format rather than on a hypothetical complete dataset [217], and the assessment should be reported with the release. This is a governance practice, not a scientific hypothesis, and we no longer describe it as a test.

7.2.3. From Single-Domain to Multi-Scenario Adaptation

Cross-cultural and cross-scenario learning requires fusion datasets that deliberately counter the WEIRD default [216,247], hybrid latent-class and Bayesian multilevel structures, and temporal networks with parameter transfer. *Recommendation*. Evaluate every model that claims generality on held-out cultures and scenarios rather than on random splits of a pooled sample [168,179]. *Testable prediction*. Hypothesis: the performance drop under a stated domain shift (a held-out culture or scenario) will be smaller for models whose person–environment coupling is represented explicitly, as in D3 mechanistic and stock–flow formulations, than for models of matched in-domain accuracy in which context enters only as a covariate. Outcome: the ratio of held-out-domain to in-domain accuracy. Criterion: a pre-registered difference in that ratio. Rejection: equal or larger drops for the explicit-coupling models.

7.2.4. Dynamic Learning and Continual Evolution

Models should update parameters as behavior and environment change, using incremental and few-shot learning together with simulation-based inference for robust nonlinear tracking [20,158]. The near-term clinical payoff is concrete, since early-warning signals of state transitions are already empirically supported [60] and pairing them with consent-governed passive sensing [93] would convert snapshot screening into trajectory monitoring. *Recommendation.* Validate any model intended to capture change on temporal generalization, not on cross-sectional fit. *Testable prediction.* Hypothesis: early-warning indicators computed from EMA or passive-sensing series at a stated sampling frequency will detect clinically defined transitions prospectively in held-out individuals. Outcome: sensitivity and false-alarm rate at a pre-specified lead time. Criterion: sensitivity above a stated floor at a false-alarm rate below a stated ceiling, both set before the data are seen. Rejection: performance at or below that of a baseline that uses the current symptom level alone; post hoc detection in already-labeled series does not count.

7.2.5. Accountable Human–AI Collaboration

A bidirectional design should let AI act as a cognitive aid while humans retain ethical review and final authority, pairing rapid recognition with oversight and real-time correction [3,183]. *Recommendation.* The evidentiary standard should rise with autonomy: randomized evaluation is the appropriate bar for any system that acts on, rather than merely describes, a person's psychological state, and the Therabot trial shows that such evaluation is feasible [195]. The informative next trial is against an active comparator rather than a waitlist, and the informative next simulation study is one that pre-registers a collective outcome and validates against held-out human data (Section 4.6). These are standards for evidence, not hypotheses, and are labeled accordingly.

8. Conclusions

Psychological and behavioral modeling has moved from theory-driven rule-based models, through data-driven machine learning, to paradigms shaped by deep learning, large language models, and foundation models of cognition. The change is more than technical iteration, since it reflects a shift from macro-level correlation to micro-level mechanism, from static inference to dynamic simulation, and from explanatory validation to predictive intervention.

Organizing the field along nested levels and five modeling classes, with system dynamics placed as a named family and as one of the aggregation operators, yields four judgments rather than a catalogue, and each is stated with its boundary. Models inherit their ontologies from psychological theories whose computational readiness, scored on formal specification and measurement maturity, varies systematically, and that gradient rather than algorithmic capability determines where theory-informed learning is currently feasible (C1, a conjecture with a stated test). Along the ladder from individual to collective units the dominant use of models shifts from estimation validated against the units' own data to simulation validated by macro-pattern reproduction, and validation practice has not followed; importing psychometric validation logic into generative simulation is the most valuable methodological transfer now available (C2, restated as a claim about use and validation rather than model form). The binding constraint on cross-scale modeling is multilevel data, because the resources that observe individuals together with their interactions are few and rarely record the collective outcome a simulation targets, so the aggregation operator and the susceptibility–exposure split remain unidentified (C3, with the required resource now specified). The evidence base is heterogeneous in quantity and quality, and readiness claims must be indexed to an explicit evidence tier (C4). Against

that background five bottlenecks recur: interpretability, privacy and ethics, cultural bias and weak generalization, insufficient and closed treatment of temporal dynamics, and the risks of one-directional human–AI decision-making. Progress will depend on integrating knowledge-driven and data-driven modeling under designs that identify causal structure rather than merely fit it, on improving dynamic and lifelong learning for long-horizon deployment, on building complementary human–AI collaboration held to trial-level evidence, and on responsible innovation and ethical governance. Pursued in this way, psychological and behavioral modeling can continue to serve as a bridge between the inner world and the external digital world.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/systems14101327/s1>. S1: The full search strings, the list of included studies with their codes, and the criterion-level readiness scores of Section 3.1 are provided as Supplementary Material. S2: The remaining included studies are listed, with their codes, in the Supplementary Material. S3: Field-wide statements about frequency (“scarce,” “few,” “most”) refer to the coded validation type and tier fields of the included set, and the corresponding counts are given in the Supplementary Material; where a statement refers to the literature beyond the included set we say so and phrase it as an impression rather than a count. S4: Maintext reference (Section 3.1): Two authors scored the fourteen families independently against the criteria; initial disagreements occurred on 3 of 28 scores, all at the boundary between adjacent formal specification levels, and were resolved by discussion, with the criterion level justifications recorded in the Supplementary Material.

Author Contributions: Conceptualization, X.C. and X.L.; methodology, X.C., X.L. and Y.P.; software, L.Y.; validation, H.Z.; formal analysis, Z.X.; writing—original draft, X.C.; writing—review and editing, X.L. and H.Z.; visualization, L.Y. and Z.X.; investigation, Z.X.; supervision, Y.P. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. The full search strings, the list of included studies with their class, evidence-tier, and validation-type codes, and the criterion-level readiness scores of Table 2 are provided as Supplementary Material. All datasets discussed in Section 5 are third-party resources; availability, licensing terms and documented biases should be confirmed with the original providers before use.

Acknowledgments: We express our gratitude to all authors for their work and dedication. We used AI tools, including GPT-5.6 and Nano Banana Pro, only for translation, language polishing, drafting of certain workflow templates, and minor formatting adjustments. All generated content has been carefully reviewed, revised, and validated by the authors. All authors take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABM	Agent-Based Modeling
ACT-R	Adaptive Control of Thought—Rational
AI	Artificial Intelligence
DDM	Drift-Diffusion Model
EMA	Ecological Momentary Assessment
LLM	Large Language Model
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT	Randomized Controlled Trial
SD	System Dynamics (Stock–Flow–Feedback Modeling)
SEM	Structural Equation Modeling
SOR	Stimulus–Organism–Response
WEIRD	Western, Educated, Industrialized, Rich, and Democratic

References

1. Song, T.; Hao, C.; Luoma, K.; Junkai, Y.; Kaiping, P. Psychoinformatics: Advances and perspectives in the computational cognition era. *J. Psychol. Sci.* **2025**, *48*, 792–803. <https://doi.org/10.16719/j.cnki.1671-6981.20250403>.
2. Cheng, N.; Fan, Y.; Liu, J.; Wang, Z.L. Application of basic emotions theory in construction of artificial psychology model. *Comput. Eng.* **2005**, *31*, 175–177.
3. Liu, F.; Wang, P.; Hu, J.; Shen, S.; Wang, H.; Shi, C.; Peng, Y.; Zhou, A. A psychologically interpretable artificial intelligence framework for the screening of loneliness, depression, and anxiety. *Appl. Psychol. Health Well-Being* **2025**, *17*, e12639. <https://doi.org/10.1111/aphw.12639>.
4. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
5. Demszky, D.; Yang, D.; Yeager, D.S.; Bryan, C.J.; Clapper, M.; Chandhok, S.; Eichstaedt, J.C.; Hecht, C.; Jamieson, J.; Johnson, M.; et al. Using large language models in psychology. *Nat. Rev. Psychol.* **2023**, *2*, 688–701. <https://doi.org/10.1038/s44159-023-00241-5>.
6. Poria, S.; Cambria, E.; Bajpai, R.; Hussain, A. A review of affective computing: From unimodal analysis to multimodal fusion. *Inf. Fusion* **2017**, *37*, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>.
7. Prinzi, F.; Barbiero, P.; Greco, C.; Amorese, T.; Cordasco, G.; Liò, P.; Vitabile, S.; Esposito, A. Using AI explainable models and handwriting/drawing tasks for psychological well-being. *Inf. Syst.* **2025**, *127*, 102465. <https://doi.org/10.1016/j.is.2024.102465>.
8. Henninger, M.; Debelak, R.; Rothacher, Y.; Strobl, C. Interpretable machine learning for psychological research: Opportunities and pitfalls. *Psychol. Methods* **2025**, *30*, 271. <https://doi.org/10.1037/met0000560>.
9. Forrester, J.W. *Industrial Dynamics*; MIT Press: Cambridge, MA, USA, 1961.
10. Stermann, J.D. *Business Dynamics: Systems Thinking and Modeling for a Complex World*; Irwin/McGraw-Hill: Boston, MA, USA, 2000.
11. Wittenborn, A.K.; Rahmandad, H.; Rick, J.; Hosseinichimeh, N. Depression as a systemic syndrome: Mapping the feedback loops of major depressive disorder. *Psychol. Med.* **2016**, *46*, 551–562. <https://doi.org/10.1017/S0033291715002044>.
12. Hosseinichimeh, N.; Rahmandad, H.; Wittenborn, A.K. Modeling the hypothalamus–pituitary–adrenal axis: A review and extension. *Math. Biosci.* **2015**, *268*, 52–65. <https://doi.org/10.1016/j.mbs.2015.08.004>.
13. Vancouver, J.B.; Weinhardt, J.M. Modeling the mind and the milieu: Computational modeling for micro-level organizational researchers. *Organ. Res. Methods* **2012**, *15*, 602–623. <https://doi.org/10.1177/1094428112449655>.
14. Homer, J.B.; Hirsch, G.B. System dynamics modeling for public health: Background and opportunities. *Am. J. Public Health* **2006**, *96*, 452–458. <https://doi.org/10.2105/AJPH.2005.062059>.
15. Rahmandad, H.; Sterman, J. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Manag. Sci.* **2008**, *54*, 998–1014. <https://doi.org/10.1287/mnsc.1070.0787>.
16. Brickman, J.; Gupta, M.; Oltmanns, J.R. Large language models for psychological assessment: A comprehensive overview. *Adv. Methods Pract. Psychol. Sci.* **2025**, *8*, 25152459251343582. <https://doi.org/10.1177/25152459251343582>.
17. Gao, C.; Lan, X.; Li, N.; Yuan, Y.; Ding, J.; Zhou, Z.; Xu, F.; Li, Y. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1259. <https://doi.org/10.1057/s41599-024-03611-3>.
18. Ziems, C.; Held, W.; Shaikh, O.; Chen, J.; Zhang, Z.; Yang, D. Can large language models transform computational social science? *Comput. Linguist.* **2024**, *50*, 237–291. https://doi.org/10.1162/coli_a_00502.
19. Wang, J.; Fang, S. The developmental trajectory of college students' psychological flexibility: Based on a latent growth model. *J. Context. Behav. Sci.* **2024**, *32*, 100765. <https://doi.org/10.1016/j.jcbs.2024.100765>.
20. Whitaker, M.M.; Bergeman, C.S.; Deboeck, P.R. The Bayesian reservoir model of psychological regulation. *Psychol. Methods* **2024**, Advance online publication. <https://doi.org/10.1037/met0000690>.
21. Carvalho, R.G.; Castro, D. Digital footprints and machine learning in psychological assessment. *Eur. Psychol.* **2026**, *31*, 274. <https://doi.org/10.1027/1016-9040/a000549>.
22. Grobelny, J.; Szymański, K.; Strozyk, Z. Act as an expert in psychometry. The evaluation of large language models utility in psychological tests cross-cultural adaptations. *Acta Psychol.* **2025**, *261*, 105813. <https://doi.org/10.1016/j.actpsy.2025.105813>.

23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, 372, n71. <https://doi.org/10.1136/bmj.n71>.
24. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, 169, 467–473. <https://doi.org/10.7326/M18-0850>.
25. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, 6, 169–200. <https://doi.org/10.1080/02699939208411068>.
26. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, 39, 1161–1178. <https://doi.org/10.1037/h0077714>.
27. Scherer, K.R. The dynamic architecture of emotion: Evidence for the component process model. *Cogn. Emot.* **2009**, 23, 1307–1351. <https://doi.org/10.1080/02699930902928969>.
28. Lazarus, R.S.; Folkman, S. *Stress, Appraisal, and Coping*; Springer: New York, NY, USA, 1984.
29. Picard, R.W. *Affective Computing*; MIT Press: Cambridge, MA, USA, 1997.
30. Goldberg, L.R. The structure of phenotypic personality traits. *Am. Psychol.* **1993**, 48, 26–34. <https://doi.org/10.1037/0003-066X.48.1.26>.
31. Roads, B.D.; Love, B.C. Modeling similarity and psychological space. *Annu. Rev. Psychol.* **2024**, 75, 215–240. <https://doi.org/10.1146/annurev-psych-040323-115131>.
32. Borsboom, D. A network theory of mental disorders. *World Psychiatry* **2017**, 16, 5–13. <https://doi.org/10.1002/wps.20375>.
33. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **1979**, 47, 263–291. <https://doi.org/10.2307/1914185>.
34. Tversky, A.; Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **1992**, 5, 297–323. <https://doi.org/10.1007/BF00122574>.
35. Ratcliff, R.; McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **2008**, 20, 873–922. <https://doi.org/10.1162/neco.2008.12-06-420>.
36. Busemeyer, J.R.; Townsend, J.T. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* **1993**, 100, 432–459. <https://doi.org/10.1037/0033-295X.100.3.432>.
37. Anderson, J.R.; Bothell, D.; Byrne, M.D.; Douglass, S.; Lebiere, C.; Qin, Y. An integrated theory of the mind. *Psychol. Rev.* **2004**, 111, 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>.
38. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
39. Ryan, R.M.; Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **2000**, 55, 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>.
40. Bandura, A. Social cognitive theory: An agentic perspective. *Annu. Rev. Psychol.* **2001**, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>.
41. Mehrabian, A.; Russell, J.A. *An Approach to Environmental Psychology*; MIT Press: Cambridge, MA, USA, 1974.
42. Helbing, D.; Molnár, P. Social force model for pedestrian dynamics. *Phys. Rev. E* **1995**, 51, 4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>.
43. Robinaugh, D.J.; Hoekstra, R.H.A.; Toner, E.R.; Borsboom, D. The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychol. Med.* **2020**, 50, 353–366. <https://doi.org/10.1017/S0033291719003404>.
44. Sijtsma, K.; van der Ark, L.A. *Measurement Models for Psychological Attributes*; CRC Press: Boca Raton, FL, USA, 2020.
45. Evans, J.S.B.T.; Stanovich, K.E. Dual-process theories of higher cognition: Advancing the debate. *Perspect. Psychol. Sci.* **2013**, 8, 223–241. <https://doi.org/10.1177/1745691612460685>.
46. Anderson, J.R.; Schooler, L.J. Reflections of the environment in memory. *Psychol. Sci.* **1991**, 2, 396–408. <https://doi.org/10.1111/j.1467-9280.1991.tb00174.x>.
47. Miyake, A.; Friedman, N.P.; Emerson, M.J.; Witzki, A.H.; Howerter, A.; Wager, T.D. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cogn. Psychol.* **2000**, 41, 49–100. <https://doi.org/10.1006/cogp.1999.0734>.
48. Shen, X.; Hu, Z.; Chen, Q.; Liu, S.; Liang, R.; Sun, J. Evolvable psychology-informed neural network for memory behavior modeling. *Inf. Process. Manag.* **2026**, 63, 104312. <https://doi.org/10.1016/j.ipm.2025.104312>.
49. Haines, N.; Kvam, P.D.; Irving, L.; Smith, C.T.; Beauchaine, T.P.; Pitt, M.A.; Ahn, W.-Y.; Turner, B.M. A tutorial on using generative models to advance psychological science: Lessons from the reliability paradox. *Psychol. Methods* **2025**, Advance online publication. <https://doi.org/10.1037/met0000674>.
50. He, C.; Peng, S.; Zhu, Y.; Wang, D.; Cai, Y.; Tu, D. Faking modeling for forced-choice measures in personality assessment based on the RES theoretical framework. *Acta Psychol. Sin.* **2025**, 57, 1832. <https://doi.org/10.3724/SP.J.1041.2025.1832>.
51. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, 3, 422–440. <https://doi.org/10.1038/s42254-021-00314-5>.

52. von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Walczak, M.; Pfrommer, J.; Pick, A.; et al. Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 614–633. <https://doi.org/10.1109/TKDE.2021.3079836>.
53. Binz, M.; Akata, E.; Bethge, M.; Brändle, F.; Callaway, F.; Coda-Forno, J.; Dayan, P.; Demircan, C.; Eckstein, M.K.; Éltető, N.; et al. A foundation model to predict and capture human cognition. *Nature* **2025**, *644*, 1002–1009. <https://doi.org/10.1038/s41586-025-09215-4>.
54. Yarkoni, T.; Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* **2017**, *12*, 1100–1122. <https://doi.org/10.1177/1745691617693393>.
55. Guest, O.; Martin, A.E. How computational modeling can force theory building in psychological science. *Perspect. Psychol. Sci.* **2021**, *16*, 789–802. <https://doi.org/10.1177/1745691620970585>.
56. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **2009**, *81*, 591–646. <https://doi.org/10.1103/RevModPhys.81.591>.
57. Centola, D. The spread of behavior in an online social network experiment. *Science* **2010**, *329*, 1194–1197. <https://doi.org/10.1126/science.1185231>.
58. Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7280–7287. <https://doi.org/10.1073/pnas.082080899>.
59. Keum, B.T.H. Conceptual application of the group actor–partner interdependence model for person–group psychological research. *Transl. Issues Psychol. Sci.* **2018**, *4*, 340. <https://doi.org/10.1037/tps0000180>.
60. van de Leemput, I.A.; Wichers, M.; Cramer, A.O.J.; Borsboom, D.; Tuerlinckx, F.; Kuppens, P.; van Nes, E.H.; Viechtbauer, W.; Giltay, E.J.; Aggen, S.H.; et al. Critical slowing down as early warning for the onset and termination of depression. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 87–92. <https://doi.org/10.1073/pnas.1312114110>.
61. Shiffman, S.; Stone, A.A.; Hufford, M.R. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* **2008**, *4*, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>.
62. Molenaar, P.C.M. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Meas. Interdiscip. Res. Perspect.* **2004**, *2*, 201–218. https://doi.org/10.1207/s15366359mea0204_1.
63. Calvo, R.A.; D’Mello, S. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* **2010**, *1*, 18–37. <https://doi.org/10.1109/T-AFFC.2010.1>.
64. Shatte, A.B.R.; Hutchinson, D.M.; Teague, S.J. Machine learning in mental health: A scoping review of methods and applications. *Psychol. Med.* **2019**, *49*, 1426–1448. <https://doi.org/10.1017/S0033291719000151>.
65. Dwyer, D.B.; Falkai, P.; Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **2018**, *14*, 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>.
66. Chekroud, A.M.; Bondar, J.; Delgadillo, J.; Doherty, G.; Wasil, A.; Fokkema, M.; Cohen, Z.; Belgrave, D.; DeRubeis, R.; Iniesta, R.; et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* **2021**, *20*, 154–170. <https://doi.org/10.1002/wps.20882>.
67. Zhang, J.; Qu, Q.; Chen, X.B. A review on collective behavior modeling and simulation: Building a link between cognitive psychology and physical action. *Appl. Intell.* **2023**, *53*, 25954–25983. <https://doi.org/10.1007/s10489-023-04924-7>.
68. Jia, S.; Zhang, Y.; Wang, F. Mapping the landscape of AI-empowered psychology: A topic modeling-based bibliometric analysis. *Artif. Intell. Rev.* **2026**, *59*, 151. <https://doi.org/10.1007/s10462-026-11543-4>.
69. Wang, M.C.; Deng, Q.; Bi, X.Y. Bayesian methods for latent-variable modeling. *Adv. Psychol. Sci.* **2017**, *25*, 1682–1695. <https://doi.org/10.3724/SP.J.1042.2017.01682>.
70. Sun, Y.; He, M.; Yang, L. Comprehensive influence model of preschool children’s personality development based on a Bayesian network. *Abstr. Appl. Anal.* **2014**, *2014*, 267953. <https://doi.org/10.1155/2014/267953>.
71. Shi, D.; Christensen, A.P.; Day, E.A.; Golino, H.F.; Garrido, L.E. Exploring estimation procedures for reducing dimensionality in psychological network modeling. *Multivar. Behav. Res.* **2025**, *60*, 184–210. <https://doi.org/10.1080/00273171.2024.2395941>.
72. Gisch, H.; Perlinger, J.; Zimmermann, J.; Kretschmar, T. The latent structure of psychodynamic and psychological personality models. *Z. Psychosom. Med. Psychother.* **2021**, *67*, 241–255. <https://doi.org/10.13109/zptm.2021.67.0a5>.
73. Navarro, D.J. Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Comput. Brain Behav.* **2019**, *2*, 28–34. <https://doi.org/10.1007/s42113-018-0019-z>.
74. Johnson, D.J.; Hopwood, C.J.; Cesario, J.; Pleskac, T.J. Advancing research on cognitive processes in social and personality psychology: A hierarchical drift diffusion model primer. *Soc. Psychol. Personal. Sci.* **2017**, *8*, 413–423. <https://doi.org/10.1177/1948550617703174>.
75. Wilson, R.C.; Collins, A.G.E. Ten simple rules for the computational modeling of behavioral data. *eLife* **2019**, *8*, e49547. <https://doi.org/10.7554/eLife.49547>.
76. Palminteri, S.; Wyart, V.; Koehlin, E. The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* **2017**, *21*, 425–433. <https://doi.org/10.1016/j.tics.2017.03.011>.

77. Bendixen, T.; Purzycki, B.G. Cognitive and cultural models in psychological science: Modeling free-list data in Bayesian regression. *Psychol. Methods* **2025**, *30*, 223. <https://doi.org/10.1037/met0000553>.
78. Hong, S.; Yang, J.H. Exploring the driver's mental control model: Concepts and insights. In Proceedings of the International Conference on Human–Computer Interaction, Copenhagen, Denmark, 23–28 July 2023; pp. 431–437. https://doi.org/10.1007/978-3-031-49215-0_51.
79. D'Mello, S.; Graesser, A. Dynamics of affective states during complex learning. *Learn. Instr.* **2012**, *22*, 145–157. <https://doi.org/10.1016/j.learninstruc.2011.10.001>.
80. Whitehill, J.; Serpell, Z.; Lin, Y.C.; Foster, A.; Movellan, J.R. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Trans. Affect. Comput.* **2014**, *5*, 86–98. <https://doi.org/10.1109/TAFFC.2014.2316163>.
81. Dubois, J.; Galdi, P.; Paul, L.K.; Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos. Trans. R. Soc. B Biol. Sci.* **2018**, *373*, 20170284. <https://doi.org/10.1098/rstb.2017.0284>.
82. Kosinski, M. Evaluating large language models in theory of mind tasks. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2405460121. <https://doi.org/10.1073/pnas.2405460121>.
83. Strachan, J.W.A.; Albergo, D.; Borghini, G.; Pansardi, O.; Scaliti, E.; Gupta, S.; Saxena, K.; Rufo, A.; Panzeri, S.; Manzi, G.; et al. Testing theory of mind in large language models and humans. *Nat. Hum. Behav.* **2024**, *8*, 1285–1295. <https://doi.org/10.1038/s41562-024-01882-z>.
84. Joel, S.; Eastwick, P.W.; Finkel, E.J. Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychol. Sci.* **2017**, *28*, 1478–1489. <https://doi.org/10.1177/0956797617714580>.
85. Peterson, J.C.; Bourgin, D.D.; Agrawal, M.; Reichman, D.; Griffiths, T.L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **2021**, *372*, 1209–1214. <https://doi.org/10.1126/science.abe2629>.
86. Lake, B.M.; Ullman, T.D.; Tenenbaum, J.B.; Gershman, S.J. Building machines that learn and think like people. *Behav. Brain Sci.* **2017**, *40*, e253. <https://doi.org/10.1017/S0140525X16001837>.
87. Kosinski, M.; Stillwell, D.; Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5802–5805. <https://doi.org/10.1073/pnas.1218772110>.
88. Youyou, W.; Kosinski, M.; Stillwell, D. Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1036–1040. <https://doi.org/10.1073/pnas.1418680112>.
89. Eichstaedt, J.C.; Smith, R.J.; Merchant, R.M.; Ungar, L.H.; Crutchley, P.; Preoțiu-Pietro, D.; Asch, D.A.; Schwartz, H.A. Facebook language predicts depression in medical records. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11203–11208. <https://doi.org/10.1073/pnas.1802331115>.
90. De Choudhury, M.; Gamon, M.; Counts, S.; Horvitz, E. Predicting depression via social media. In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), Cambridge, MA USA, 8–11 July 2013; Volume 7, pp. 128–137. <https://doi.org/10.1609/icwsml.v7i1.14432>.
91. Guntuku, S.C.; Yaden, D.B.; Kern, M.L.; Ungar, L.H.; Eichstaedt, J.C. Detecting depression and mental illness on social media: An integrative review. *Curr. Opin. Behav. Sci.* **2017**, *18*, 43–49. <https://doi.org/10.1016/j.cobeha.2017.07.005>.
92. Onnela, J.P.; Rauch, S.L. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* **2016**, *41*, 1691–1696. <https://doi.org/10.1038/npp.2016.7>.
93. Mohr, D.C.; Zhang, M.; Schueller, S.M. Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annu. Rev. Clin. Psychol.* **2017**, *13*, 23–47. <https://doi.org/10.1146/annurev-clinpsy-032816-044949>.
94. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Personality predictions based on user behavior on the Facebook social-media platform. *IEEE Access* **2018**, *6*, 61959–61969. <https://doi.org/10.1109/ACCESS.2018.2876502>.
95. Choi, J.H.; Jung, D.H. The role of psychological factors in predicting self-rated health: Implications from machine-learning models. *Psychol. Health Med.* **2025**, *30*, 1158–1170. <https://doi.org/10.1080/13548506.2025.2450546>.
96. Pereira, M.G.; Santos, M.; Magalhães, R.; Rodrigues, C.; Araújo, O.; Durães, D. Burnout risk profiles in psychology students: An exploratory study with machine learning. *Behav. Sci.* **2025**, *15*, 505. <https://doi.org/10.3390/bs15040505>.
97. Dong, J.; Chu, M.; Xu, Y. A predictive healthcare model using machine learning and psychological factors for medication adherence. *Healthc. Anal.* **2025**, *7*, 100397. <https://doi.org/10.1016/j.health.2025.100397>.
98. Kumuda, A.; Panigrahy, S.K. Design of an efficient model for psychological-disease analysis and prediction using machine learning and genomic data. *Big Data Cogn. Comput.* **2025**, *9*, 49. <https://doi.org/10.3390/bdcc9030049>.
99. Zhou, Y.; Zhang, Z.; Xiang, F.; Xu, M.; Zhou, Z. Assessing psychological resilience and its influencing factors in the MSM population by machine learning. *Sci. Rep.* **2025**, *15*, 24022. <https://doi.org/10.1038/s41598-025-08551-9>.
100. Fumero, A.; Navarrete, G. Personality and psychological distress: Application of a neural network model. *Rev. Iberoam. Diagn. Eval.* **2016**, *41*, 28–38.
101. Li, N.; Zhai, H.; Seetharam, T.G.; Shanthini, A. Psychological health analysis based on a fuzzy-assisted neural-network model for sportspersons. *J. Intell. Fuzzy Syst.* **2021**, 1–15. <https://api.semanticscholar.org/CorpusID:236665790> (accessed on 9 October 2026).

102. Zhou, X.; Liang, Z.; Zhang, G. Using explainable machine learning to investigate the relationship between childhood maltreatment, positive psychological traits, and CPTSD symptoms. *Eur. J. Psychotraumatol.* **2025**, *16*, 2455800. <https://doi.org/10.1080/20008066.2025.2455800>.
103. Cohen, I.; Garg, A.; Huang, T.S. Emotion recognition from facial expressions using multilevel HMM. In Proceedings of the NeurIPS, Breckenridge, CO, USA, 1–2 December 2000.
104. Schuller, B.; Rigoll, G.; Lang, M. Hidden Markov model-based speech emotion recognition. In Proceedings of the IEEE ICASSP, Hong Kong, 6–10 April 2003. <https://doi.org/10.1109/ICASSP.2003.1202279>.
105. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. <https://doi.org/10.1109/TPAMI.2018.2798607>.
106. Poria, S.; Majumder, N.; Mihalcea, R.; Hovy, E. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access* **2019**, *7*, 100943–100953. <https://doi.org/10.1109/ACCESS.2019.2929050>.
107. Toichoa Eyam, A.; Mohammed, W.M.; Martinez Lastra, J.L. Emotion-driven analysis and control of human–robot interactions in collaborative applications. *Sensors* **2021**, *21*, 4626. <https://doi.org/10.3390/s21144626>.
108. Almqren, M.A.; Almqren, L.; Alhayan, F.; Cristea, A.I.; Pennington, D. Using deep learning to analyze the psychological effects of COVID-19. *Front. Psychol.* **2023**, *14*, 962854. <https://doi.org/10.3389/fpsyg.2023.962854>.
109. Alqazzaz, A.; Quasim, M.T.; Alshahrani, M.M.; Alrashdi, I.; Khan, M.A. A deep learning model to analyse social-cyber psychological problems in youth. *Comput. Syst. Sci. Eng.* **2023**, *46*, 551–562. <https://doi.org/10.32604/csse.2023.031048>.
110. Acciai, A.; Guerrisi, L.; Plebe, A.; Suriano, R. Neural language models as content-analysis tools in psychology. *Philos. Psychol.* **2025**, *39*, 2519–2542. <https://doi.org/10.1080/09515089.2025.2508942>.
111. Binz, M.; Schulz, E. Using cognitive psychology to understand GPT-3. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2218523120. <https://doi.org/10.1073/pnas.2218523120>.
112. Bhandari, P.; Naseem, U.; Datta, A.; Fay, N.; Nasim, M. Evaluating personality traits in large language models: Insights from psychological questionnaires. In Proceedings of the Companion Proceedings of the ACM Web Conference, Sydney, NSW, Australia, 28 April–2 May 2025; pp. 868–872. <https://doi.org/10.1145/3701716.3715504>.
113. Pellert, M.; Lechner, C.M.; Wagner, C.; Rammstedt, B.; Strohmaier, M. AI psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspect. Psychol. Sci.* **2024**, *19*, 808–826. <https://doi.org/10.1177/17456916231214460>.
114. Santurkar, S.; Durmus, E.; Ladhak, F.; Lee, C.; Liang, P.; Hashimoto, T. Whose opinions do language models reflect? In Proceedings of the 40th International Conference on Machine Learning (ICML), Honolulu, HI, USA, 23–29 July 2023; pp. 29971–30004.
115. Tjautja, L.; Chen, V.; Wu, T.; Talwalkar, A.; Neubig, G. Do LLMs exhibit human-like response biases? A case study in survey design. *Trans. Assoc. Comput. Linguist.* **2024**, *12*, 1011–1026. https://doi.org/10.1162/tacl_a_00685.
116. IJntema, R.C.; Schaufeli, W.B.; Burger, Y.D. Resilience mechanisms at work: The psychological immunity–psychological elasticity (PI-PE) model. *Curr. Psychol.* **2023**, *42*, 4719–4731. <https://doi.org/10.1007/s12144-021-01813-5>.
117. Zewude, G.T.; Mesfin, Y.; Sadouki, F.; Ayele, A.G.; Goraw, S.; Segon, T.; Hercz, M. A serial mediation model of Big 5 personality traits, emotional intelligence, and psychological capital as predictors of teachers’ professional well-being. *Acta Psychol.* **2024**, *250*, 104500. <https://doi.org/10.1016/j.actpsy.2024.104500>.
118. Kaçar-Başaran, S. From hero to zero: A serial mediation model between narcissistic perfectionism and psychological distress. *J. Ration.-Emotive Cogn.-Behav. Ther.* **2025**, *43*, 4. <https://doi.org/10.1007/s10942-024-00572-z>.
119. Hu, Y.; Fei, J.; Yue, J.; Gao, R.; Song, Q.; Zhao, X.; Mei, S. Emotional abuse and neglect, and depression: A moderated mediation model. *Psychiatry Investig.* **2025**, *22*, 389–396. <https://doi.org/10.30773/pi.2024.0108>.
120. Barnes, C.; Mongrain, M. A three-factor model of personality predicts changes in depression and subjective well-being following positive-psychology interventions. *J. Posit. Psychol.* **2020**, *15*, 556–568. <https://doi.org/10.1080/17439760.2019.1651891>.
121. Wierds, C.M.; Faulkner, G.; Rhodes, R.E.; Zumbo, B.D.; Beauchamp, M.R. Psychological needs and exercise behaviour: A comparison of two psychological needs models. *Psychol. Health* **2024**, *39*, 1284–1302. <https://doi.org/10.1080/08870446.2022.2141484>.
122. Li, J.; Pan, X. The impact of mood on sports flow state in football players: A chain mediating model. *Front. Psychol.* **2025**, *16*, 1523400. <https://doi.org/10.3389/fpsyg.2025.1523400>.
123. Stricker, J.; Pietrowsky, R. Incremental validity of the ICD-11 personality disorder model for explaining psychological distress. *Personal. Disord. Theory Res. Treat.* **2022**, *13*, 97–107. <https://doi.org/10.1037/per0000489>.
124. Shafaei, A.; Nejati, M.; Omari, M.; Sharafizad, F. Inclusive leadership and workplace bullying: A model of psychological safety, self-esteem, and embeddedness. *J. Leadersh. Organ. Stud.* **2024**, *31*, 41–58. <https://doi.org/10.1177/15480518231209018>.
125. Soltani Nejad, N.; Taheri Azad, F.; Ashraf Ganjouie, A.; Kooshan, Z.; Rastegar, R. Job embeddedness in the hospitality industry: A model of psychological antecedents and behavioral consequences. *Int. J. Hosp. Tour. Adm.* **2025**, *26*, 160–179. <https://doi.org/10.1080/15256480.2023.2276959>.

126. Ring, M.; Hult, M. A structural equation model of the impacts of nurses' psychological safety and psychological contract breach. *J. Adv. Nurs.* **2025**, *81*, 1323–1331. <https://doi.org/10.1111/jan.16331>.
127. Zhang, J.; Choi, M.C.; Wang, K.; Kim, H.E. How can ethical leadership increase employees' bootlegging innovation behavior in China? *Front. Psychol.* **2025**, *16*, 1506906. <https://doi.org/10.3389/fpsyg.2025.1506906>.
128. Ren, X.; Xu, S.; Yu, J.; Lin, F. The impact of project managers' psychological resilience on adaptive performance: A multilevel model. *Proj. Manag. J.* **2025**, *56*, 198–213. <https://doi.org/10.1177/87569728241262222>.
129. Newaz, M.T.; Davis, P.; Jefferies, M.; Pillay, M. The psychological contract: A missing link between safety climate and safety behaviour on construction sites. *Saf. Sci.* **2019**, *112*, 9–17. <https://doi.org/10.1016/j.ssci.2018.10.002>.
130. Bottaro, R.; De Giovanni, K.; Faraci, P. Technostress and work–life fit for employees' psychological well-being: A cross-cultural mediation model. *Psychol. Rep.* **2025**, *17*, 00332941251320556. <https://doi.org/10.1177/00332941251320556>.
131. Arpaci, I.; Bahari, M. Investigating the role of psychological needs in predicting the educational sustainability of Metaverse using a deep learning-based hybrid SEM-ANN technique. *Interact. Learn. Environ.* **2024**, *32*, 2957–2969. <https://doi.org/10.1080/10494820.2022.2164313>.
132. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009. <https://doi.org/10.1017/CBO9780511803161>.
133. Rohrer, J.M. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Adv. Methods Pract. Psychol. Sci.* **2018**, *1*, 27–42. <https://doi.org/10.1177/2515245917745629>.
134. Hamaker, E.L.; Kuiper, R.M.; Grasman, R.P.P. A critique of the cross-lagged panel model. *Psychol. Methods* **2015**, *20*, 102–116. <https://doi.org/10.1037/a0038889>.
135. Jia, X.; Wei, X. Student behavioral temporal modeling sensitive to abnormal behavior for mental-health prediction. *J. Beijing Univ. Technol.* **2024**, *50*, 939–947. <https://doi.org/10.11936/bjutxb2023010010>.
136. Privado, J.; Pérez-Eizaguirre, M.; Martínez-Rodríguez, M.; Ponce-de León, L. Cognitive and non-cognitive factors as predictors of academic performance. *Learn. Individ. Differ.* **2024**, *116*, 102536. <https://doi.org/10.1016/j.lindif.2024.102536>.
137. Song, S.Y.; Qian, K. A study on the effect of deep reinforcement learning in cultivating athlete decision behavior and psychological resilience. *Scalable Comput. Pract. Exp.* **2025**, *26*, 250–258.
138. Park, J.S.; O'Brien, J.C.; Cai, C.J.; Morris, M.R.; Liang, P.; Bernstein, M.S. Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST), San Francisco, CA, USA, 29 October–1 November 2023. <https://doi.org/10.1145/3586183.3606763>.
139. Aher, G.V.; Arriaga, R.I.; Kalai, A.T. Using large language models to simulate multiple humans and replicate human subject studies. In Proceedings of the International Conference on Machine Learning (ICML), Honolulu, HI, USA, 23–29 July 2023; pp. 337–371.
140. Park, J.S.; Zou, C.Q.; Shaw, A.; Hill, B.M.; Cai, C.; Morris, M.R.; Willer, R.; Liang, P.; Bernstein, M.S. Generative agent simulations of 1,000 people. *arXiv* **2024**, arXiv:2411.10109. <https://doi.org/10.48550/arXiv.2411.10109>.
141. Argyle, L.P.; Busby, E.C.; Fulda, N.; Gubler, J.R.; Rytting, C.; Wingate, D. Out of one, many: Using language models to simulate human samples. *Political Anal.* **2023**, *31*, 337–351. <https://doi.org/10.1017/pan.2023.2>.
142. Wang, A.; Morgenstern, J.; Dickerson, J.P. Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nat. Mach. Intell.* **2025**, *7*, 400–411. <https://doi.org/10.1038/s42256-025-00986-z>.
143. Bansal, R.; Heidler, P.; Aljuaid, M.; Sakib, M.N. The psychology of retention in sports: Understanding work–life balance through the SOR model. *Rev. Psicol. Deporte* **2025**, *34*, 185–198.
144. Ozanne, L.K.; Prayag, G.; Chowdhury, M. Social capital and its psychological consequences on consumer resilience: A mediated-moderation model. *J. Consum. Aff.* **2025**, *59*, e70013. <https://doi.org/10.1111/joca.70013>.
145. Ye, Z.; Huang, X. Adoption of a deep learning-based neural network model in the psychological behavior analysis of resident tourism consumption. *Front. Public Health* **2022**, *10*, 995828. <https://doi.org/10.3389/fpubh.2022.995828>.
146. Yadav, A.; Parida, M.; Choudhary, P.; Kumar, B. Modelling traffic-noise-induced annoyance at intersections and its association with psychological health. *Transp. Res. Part D* **2025**, *139*, 104568. <https://doi.org/10.1016/j.trd.2024.104568>.
147. Tang, R.; Lu, G.; Liu, M.; Zhu, M.; Li, P. A unified driving-behavior model based on psychological safety space. *Transp. Res. Part F* **2025**, *109*, 439–457. <https://doi.org/10.1016/j.trf.2024.12.024>.
148. Li, Z.; Wu, C.; Tao, P.; Tian, J.; Ma, L. DP and DS-LCD: A new lane-change decision model coupling driver's psychology and driving style. *IEEE Access* **2020**, *8*, 132614–132624. <https://doi.org/10.1109/ACCESS.2020.3010409>.
149. Huang, J.; Liu, T.; Hu, Y.; He, Z.; Hu, L. Assessment and recognition of driver situation awareness in conditional autonomous driving: Integrating cognitive psychology and machine learning. *J. Ind. Inf. Integr.* **2025**, *47*, 100894. <https://doi.org/10.1016/j.jii.2025.100894>.
150. Brand, M.; Young, K.S.; Laier, C.; Wölfling, K.; Potenza, M.N. Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An interaction of person-affect-cognition-execution (I-PACE) model. *Neurosci. Biobehav. Rev.* **2016**, *71*, 252–266. <https://doi.org/10.1016/j.neubiorev.2016.08.033>.

151. Ren, J.; Ye, K.T.; Li, F.Y.; Chen, S. Information aggregation and psychological risk dual-driven venture capital decision: A model based on extended fuzzy sets. *Financ. Res. Lett.* **2025**, *86*, 108596. <https://doi.org/10.1016/j.frl.2025.108596>.
152. Wu, P.; Xie, Y.; Zhou, L.; Deveci, M.; Martínez, L. Multi-source data-driven decision-support model for product ranking with consumer psychology behavior. *Inf. Fusion* **2025**, *118*, 103014. <https://doi.org/10.1016/j.inffus.2025.103014>.
153. Wang, Q.; Wang, J.; Zhang, S.; Yu, F.; De Allegri, M. Psychological constructs and preferences for a complementary inclusive health insurance: A hybrid choice model. *Health Policy Plan.* **2025**, *40*, 981–991. <https://doi.org/10.1093/heapol/czaf056>.
154. Chen, J.; Yang, X. Fuzzy asset-liability portfolio optimization model with investors' mental accounts. *J. Guangdong Univ. Technol.* **2025**, *42*. <https://doi.org/10.12052/gdutxb.230178>.
155. Yin, X.; Xu, X.; Pan, B. Selection of strategy for large-group emergency decision-making based on risk measurement. *Reliab. Eng. Syst. Saf.* **2021**, *208*, 107325. <https://doi.org/10.1016/j.res.2020.107325>.
156. Oliveira, A. A discussion of rational and psychological decision-making theories and models: The search for a cultural-ethical decision-making model. *Electron. J. Bus. Ethics Organ. Stud.* **2007**, *12*, 12–17.
157. Li, Y.; Williams, L.; Muth, C.; Heshmati, S.; Chow, S.-M.; Oravecz, Z. A growth of hierarchical autoregression model for capturing individual differences in dynamic characteristics of psychological processes. *Struct. Equ. Model.* **2025**, *32*, 237–250. <https://doi.org/10.1080/10705511.2024.2402328>.
158. Wanke, P.; Chuanpeng, H. Neural simulation-based inference: A neural network and simulation-based inference approach to cognitive modelling. *J. Psychol. Sci.* **2025**, *48*, 826–835. <https://doi.org/10.16719/j.cnki.1671-6981.20250406>.
159. Zhang, Z.; Liu, Y.; Gao, M.; Yang, Y.; Cao, J.; Hou, W.K.; Li, S.; Yau, S.; Wing, Y.K.; Lee, T. Data-efficient model for psychological resilience prediction based on neurological data. *arXiv* **2025**, arXiv:2502.01377. <https://doi.org/10.48550/arXiv.2502.01377>.
160. Czeszumski, A.; Eustergerling, S.; Lang, A.; Menrath, D.; Gerstenberger, M.; Schuberth, S.; Schreiber, F.; Rendon, Z.Z.; König, P. Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.* **2020**, *14*, 39. <https://doi.org/10.3389/fnhum.2020.00039>.
161. Sanchez-Cortes, D.; Aran, O.; Schmid Mast, M.; Gatica-Perez, D. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Trans. Multimed.* **2012**, *14*, 816–832. <https://doi.org/10.1109/TMM.2011.2181941>.
162. Park, C.Y.; Cha, N.; Kang, S.; Kim, A.; Khandoker, A.H.; Hadjileontiadis, L.; Oh, A.; Jeong, Y.; Lee, U. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Sci. Data* **2020**, *7*, 293. <https://doi.org/10.1038/s41597-020-00630-y>.
163. Wang, X.; Wang, Y.; Zhang, Z. Cultivation model of football training group cohesion from the perspective of group psychology. *Entertain. Comput.* **2024**, *49*, 100626. <https://doi.org/10.1016/j.entcom.2023.100626>.
164. Kececi, O.; Talaghir, L.-G.; Celik, V.O.; Iconomescu, T.M.; Koc, M.C.; Manolache, G.M. A model on leader–member exchange, psychological empowerment and teamwork in esports teams. *Front. Psychol.* **2025**, *16*, 1653752. <https://doi.org/10.3389/fpsyg.2025.1653752>.
165. Helbing, D.; Farkas, I.; Vicsek, T. Simulating dynamical features of escape panic. *Nature* **2000**, *407*, 487–490. <https://doi.org/10.1038/35035023>.
166. Moussaïd, M.; Helbing, D.; Theraulaz, G. How simple rules determine pedestrian behavior and crowd disasters. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6884–6888. <https://doi.org/10.1073/pnas.1016507108>.
167. Zhang, W.; Zhang, Z.; Ma, Y.; Lee, E.W.M.; Shi, M. Psychological impatience in pedestrian evacuation: Modelling, simulations and experiments. *Physica A* **2024**, *635*, 129472. <https://doi.org/10.1016/j.physa.2023.129472>.
168. Wang, L.; Zhang, Z.; Lu, S.; Wang, J. Behavior-selection models of fire evacuations considering adaptive evacuation psychologies. *Sustainability* **2024**, *16*, 3607. <https://doi.org/10.3390/su16093607>.
169. Deng, K.; Li, M.; Hu, X.; Chen, T. An extended social-force model on unidirectional flow considering the impacts of a hazard source. *Chin. Phys. B* **2024**, *33*, 028901. <https://doi.org/10.1088/1674-1056/ad1173>.
170. Qiu, H.; Zhang, W.; Shi, M.; Lee, E.W.M. An improved multi-velocity cellular-automaton model that considers psychological impatience. *J. Saf. Sci. Resil.* **2025**, *6*, 100196. <https://doi.org/10.1016/j.jnlssr.2024.11.006>.
171. Wang, H.; Xu, T.; Li, F. A novel emergency evacuation model of subway station passengers considering personality traits. *Sustainability* **2021**, *13*, 10463. <https://doi.org/10.3390/su131810463>.
172. Deffuant, G.; Neau, D.; Amblard, F.; Weisbuch, G. Mixing beliefs among interacting agents. *Adv. Complex Syst.* **2000**, *3*, 87–98. <https://doi.org/10.1142/S0219525900000078>.
173. Li, Y.; Liu, M.; Cao, J.; Wang, X.; Zhang, N. Multi-attribute group decision-making considering opinion dynamics. *Expert Syst. Appl.* **2021**, *184*, 115479. <https://doi.org/10.1016/j.eswa.2021.115479>.
174. Chen, H.; Jin, J.; Sun, S.; Zhang, C.; Yu, M. Propagation dynamics of emotional contagion using an S3EIR model based on multiple social-media platforms. *Front. Commun.* **2025**, *10*, 1582974. <https://doi.org/10.3389/fcomm.2025.1582974>.
175. Wang, X.; Wang, X.; Yang, W. A rumor propagation model integrated with psychological factors. *Int. J. Biomath.* **2025**, 2450149. <https://doi.org/10.1142/s1793524524501493>.

176. Li, Y.; Zhang, Y.; Lin, J. Message-spreading modeling from the perspective of social psychology, differential-equation dynamics, and deep learning. *Nonlinear Anal. Model. Control* **2025**, *30*, 592–605. <https://doi.org/10.15388/namc.2025.30.41797>.
177. Ferrara, E.; Yang, Z. Measuring emotional contagion in social media. *PLoS ONE* **2015**, *10*, e0142390. <https://doi.org/10.1371/journal.pone.0142390>.
178. Saha, S.; Pal, S.; Melnik, R. Nonlocal cooperative behavior, psychological effects, and collective decision-making. *Math. Methods Appl. Sci.* **2025**, *48*, 12011–12037. <https://doi.org/10.1002/mma.11010>.
179. Xu, Y.; Yan, S.; Li, Y. A multi-attribute quantum group-consensus model considering psychological preference. *Eng. Appl. Artif. Intell.* **2025**, *144*, 110086. <https://doi.org/10.1016/j.engappai.2025.110086>.
180. Chen, Y.; Zhou, M.; Liu, F. A consensus model with non-cooperative behavior adaptive management based on cognitive psychological-state computation. *Cogn. Comput.* **2024**, *16*, 3344–3363. <https://doi.org/10.1007/s12559-024-10330-z>.
181. Jung, J.; Miller, J.H.; Page, S.E. Agent-based modeling for psychological research on social phenomena. *Am. Psychol.* **2025**, *81*, 1071–1084. <https://doi.org/10.1037/amp0001530>.
182. Pirla, S. A psychological model of collective risk perceptions. *Judgm. Decis. Mak.* **2024**, *19*, e13. <https://doi.org/10.1017/jdm.2024.9>.
183. Bail, C.A. Can generative AI improve social science? *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2314021121. <https://doi.org/10.1073/pnas.2314021121>.
184. Grossmann, I.; Feinberg, M.; Parker, D.C.; Christakis, N.A.; Tetlock, P.E.; Cunningham, W.A. AI and the transformation of social science research. *Science* **2023**, *380*, 1108–1109. <https://doi.org/10.1126/science.adi1778>.
185. Dillion, D.; Tandon, N.; Gu, Y.; Gray, K. Can AI language models replace human participants? *Trends Cogn. Sci.* **2023**, *27*, 597–600. <https://doi.org/10.1016/j.tics.2023.04.008>.
186. Luoma, K.; Zengyi, L.; Jiangqun, L.; Song, T.; Kaiping, P. Effectiveness of large language models in simulating regional psychological structures. *arXiv* **2025**, arXiv:2509.25283. <https://doi.org/10.48550/arXiv.2509.25283>.
187. Bak-Coleman, J.B.; Alfano, M.; Barfuss, W.; Bergstrom, C.T.; Centeno, M.A.; Couzin, I.D.; Donges, J.F.; Galesic, M.; Gersick, A.S.; Jacquet, J.; et al. Stewardship of global collective behavior. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2025764118. <https://doi.org/10.1073/pnas.2025764118>.
188. Wang, R.; Chen, F.; Chen, Z.; Li, T.; Harari, G.; Tignor, S.; Zhou, X.; Ben-Zeev, D.; Campbell, A.T. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), Seattle, WA, USA, 13–17 September 2014; pp. 3–14. <https://doi.org/10.1145/2632048.2632054>.
189. Sapiezynski, P.; Stopczynski, A.; Lassen, D.D.; Lehmann, S. Interaction data from the Copenhagen Networks Study. *Sci. Data* **2019**, *6*, 315. <https://doi.org/10.1038/s41597-019-0325-x>.
190. Cattuto, C.; Van den Broeck, W.; Barrat, A.; Colizza, V.; Pinton, J.F.; Vespignani, A. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* **2010**, *5*, e11596. <https://doi.org/10.1371/journal.pone.0011596>.
191. Eagle, N.; Pentland, A. Reality mining: Sensing complex social systems. *Pers. Ubiquitous Comput.* **2006**, *10*, 255–268. <https://doi.org/10.1007/s00779-005-0046-3>.
192. Olguín Olguín, D.; Waber, B.N.; Kim, T.; Mohan, A.; Ara, K.; Pentland, A. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2009**, *39*, 43–55. <https://doi.org/10.1109/TSMCB.2008.2006638>.
193. Carletta, J. Unleashing the killer corpus: Experiences in creating the multi-everything AMI Meeting Corpus. *Lang. Resour. Eval.* **2007**, *41*, 181–190. <https://doi.org/10.1007/s10579-007-9040-x>.
194. Busso, C.; Bulut, M.; Lee, C.-C.; Kazemzadeh, A.; Mower, E.; Kim, S.; Chang, J.N.; Lee, S.; Narayanan, S.S. IEMOCAP: Interactive emotional dyadic motion capture database. *Lang. Resour. Eval.* **2008**, *42*, 335–359. <https://doi.org/10.1007/s10579-008-9076-6>.
195. Heinz, M.V.; Mackin, D.M.; Trudeau, B.M.; Bhattacharya, S.; Wang, Y.; Banta, H.A.; Jewett, A.D.; Salzhauer, A.J.; Griffin, T.Z.; Jacobson, N.C. Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI* **2025**, *2*, AIoa2400802. <https://doi.org/10.1056/AIoa2400802>.
196. LeGris, S.; Vong, W.K.; Lake, B.M.; Gureckis, T.M. H-ARC: A robust estimate of human performance on the Abstraction and Reasoning Corpus benchmark. *arXiv* **2024**, arXiv:2409.01374. <https://doi.org/10.48550/arXiv.2409.01374>.
197. Koelstra, S.; Muehl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. <https://doi.org/10.1109/T-AFFC.2011.15>.
198. Zheng, W.L.; Lu, B.L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. <https://doi.org/10.1109/TAMD.2015.2431497>.
199. Chen, J.; Wang, X.; Huang, C.; Hu, X.; Shen, X.; Zhang, D. A large finer-grained affective computing EEG dataset. *Sci. Data* **2023**, *10*, 740. <https://doi.org/10.1038/s41597-023-02650-w>.
200. Katsigiannis, S.; Ramzan, N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 98–107. <https://doi.org/10.1109/JBHI.2017.2688239>.

201. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.* **2019**, *10*, 18–31. <https://doi.org/10.1109/TAFFC.2017.2740923>.
202. Livingstone, S.R.; Russo, F.A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE* **2018**, *13*, e0196391. <https://doi.org/10.1371/journal.pone.0196391>.
203. Zadeh, A.B.; Liang, P.P.; Poria, S.; Cambria, E.; Morency, L.P. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 2236–2246. <https://doi.org/10.18653/v1/P18-1208>.
204. Gjurković, M.; Karan, M.; Vukojević, I.; Bošnjak, M.; Šnajder, J. PANDORA talks: Personality and demographics on Reddit. In Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media, Online, 10 June 2021; pp. 138–152. <https://doi.org/10.18653/v1/2021.socialnlp-1.12>.
205. Gjurković, M.; Šnajder, J. Reddit: A gold mine for personality prediction. In Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media, New Orleans, LA, USA, 6 June 2018; pp. 87–97. <https://doi.org/10.18653/v1/W18-1112>.
206. Plank, B.; Hovy, D. Personality traits on Twitter—or—How to get 1,500 personality tests in a week. In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), Lisbon, Portugal, 17 September 2015; pp. 92–98. <https://doi.org/10.18653/v1/W15-2913>.
207. Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. Predicting personality from Twitter. In Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011; pp. 149–156. <https://doi.org/10.1109/PASSAT/SocialCom.2011.33>.
208. Gratch, J.; Artstein, R.; Lucas, G.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; DeVault, D.; Marsella, S.; et al. The Distress Analysis Interview Corpus of human and computer interviews. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), Reykjavik, Iceland, 26–31 May 2014; pp. 3123–3128.
209. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI), Boulder, CO, USA, 16–20 October 2018; pp. 400–408. <https://doi.org/10.1145/3242969.3242985>.
210. Haerpfer, C.; Inglehart, R.; Moreno, A.; Welzel, C.; Kizilova, K.; Diez-Medrano, J.; Lagos, M.; Norris, P.; Ponarin, E.; Puranen, B. *World Values Survey: Round Seven—Country-Pooled Datafile Version 5.0*; JD Systems Institute and WVSA Secretariat: Madrid, Spain; Vienna, Austria, 2022. <https://doi.org/10.14281/18241.20>.
211. European Social Survey ERIC. *European Social Survey (ESS), Rounds 1–11, Integrated Data Files*; Sikt—Norwegian Agency for Shared Services in Education and Research: Bergen, Norway, 2023.
212. Bian, Y.; Li, L. The Chinese General Social Survey (2003–8): Sample designs and data evaluation. *Chin. Sociol. Rev.* **2012**, *45*, 70–97. <https://doi.org/10.2753/CSA2162-0555450104>.
213. EVS. European Values Study 2017: Integrated Dataset (EVS 2017). GESIS Data Archive, Cologne, ZA7500 Data File Version 5.0.0. 2022. Available online: https://search.gesis.org/research_data/ZA7500 (accessed on 9 August 2026). <https://doi.org/10.4232/1.13897>.
214. Helliwell, J.F.; Layard, R.; Sasch, J.D.; De Neve, J.; Aknin, L.B.; Wang, S. *World Happiness Report 2024*; University of Oxford, Wellbeing Research Centre: Oxford, UK, 2024.
215. Human Relations Area Files. *eHRAF World Cultures Database*; Yale University: New Haven, CT, USA, 2024.
216. Henrich, J.; Heine, S.J.; Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **2010**, *33*, 61–83. <https://doi.org/10.1017/S0140525X0999152X>.
217. Rocher, L.; Hendrickx, J.M.; de Montjoye, Y.A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat. Commun.* **2019**, *10*, 3069. <https://doi.org/10.1038/s41467-019-10933-3>.
218. Prykhodko, I. The model of psychological safety of a soldier’s personality. *Curr. Issues Personal. Psychol.* **2021**, *10*, 112–122. <https://doi.org/10.5114/cipp.2021.108684>.
219. Straud, C.L.; Buccellato, K.H.; Vacek, S.; Hale, W.J.; Baker, M.T.; Isler, W.C.; Litz, B.T.; McNally, R.J.; Peterson, A.L. The General Military Support Scale: An examination of factor structure and psychological correlates. *Mil. Psychol.* **2025**, *38*, 277–228. <https://doi.org/10.1080/08995605.2025.2495371>.
220. Aygün, İ.; Kaya, B.; Kaya, M. Identifying patients in need of psychological treatment with language-representation models. *Multimed. Tools Appl.* **2025**, *84*, 397–418. <https://doi.org/10.1007/s11042-024-18992-5>.
221. Hossain, M.; Sanjara; Hossain, S.; Chaki, S.; Rahman, S.; Ali, A.B.M.S. Revolutionizing mental-health sentiment analysis with BERT-Fuse: A hybrid deep-learning model. *IEEE Access* **2025**, *13*, 85428–85446. <https://doi.org/10.1109/ACCESS.2025.3568340>.
222. Hemalatha, S. Leveraging deep learning for depression diagnosis in women: A hybrid CNN-GRU-MLP fusion model. *Int. J. Intell. Eng. Syst.* **2025**, *18*, 602–621. <https://doi.org/10.22266/ijies2025.0531.39>.

223. Brown, S.A.; Holland, J.; Gaynor, K.; Bramham, J.; O'Keeffe, F.; O'Flanagan, S.; Savinelli, S.; Mallon, P.; Feeney, E.; Kenny, G.; et al. A psychological model of predictive factors of distress following long COVID. *J. Affect. Disord.* **2025**, *373*, 394–402. <https://doi.org/10.1016/j.jad.2024.12.049>.
224. Peng, F.; Nie, J. Psychological counseling ability of large language models. *arXiv* **2025**, arXiv:2503.07627. <https://doi.org/10.48550/arXiv.2503.07627>.
225. Gao, F.; He, Y.; Chen, Q.; Liu, F. Evaluating psychological competency via Chinese Q&A in large language models. *Appl. Sci.* **2025**, *15*, 9089. <https://doi.org/10.3390/app15169089>.
226. Abbasi, M.A.; Mirnezami, F.S.; Naderi, H. PsychoLex: Unveiling the psychological mind of large language models. In Proceedings of the 1st Workshop on NLP for Science (NLP4Science), Miami, FL, USA, 12–16 November 2024; pp. 24–35. <https://doi.org/10.18653/v1/2024.nlp4science-1.4>.
227. Hu, J.; Dong, T.; Luo, G.; Ma, H.; Zou, P.; Sun, X.; Guo, D.; Yang, X.; Wang, M. PsychoLLM: Enhancing LLMs for psychological understanding and evaluation. *IEEE Trans. Comput. Soc. Syst.* **2024**, *12*, 539–551. <https://doi.org/10.1109/TCSS.2024.3497725>.
228. Lai, T.; Shi, Y.; Du, Z.; Wu, J.; Fu, K.; Dou, Y.; Wang, Z. Psy-LLM: Scaling up global mental health psychological services with AI-based large language models. *arXiv* **2023**, arXiv:2307.11991. <https://doi.org/10.48550/arXiv.2307.11991>.
229. Yang, K.; Zhang, T.; Kuang, Z.; Xie, Q.; Huang, J.; Ananiadou, S. MentaLLaMA: Interpretable mental health analysis on social media with large language models. In Proceedings of the ACM Web Conference 2024 (WWW), Singapore, 13–17 May 2024; pp. 4489–4500. <https://doi.org/10.1145/3589334.3648137>.
230. Xu, X.; Yao, B.; Dong, Y.; Gabriel, S.; Yu, H.; Hendler, J.; Ghassemi, M.; Dey, A.K.; Wang, D. Mental-LLM: Leveraging large language models for mental health prediction via online text data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2024**, *8*, 1–32. <https://doi.org/10.1145/3643540>.
231. Stadel, E.C.; Stirman, S.W.; Ungar, L.H.; Boland, C.L.; Schwartz, H.A.; Yaden, D.B.; Sedoc, J.; DeRubeis, R.J.; Willer, R.; Eichstaedt, J.C. Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *npj Ment. Health Res.* **2024**, *3*, 12. <https://doi.org/10.1038/s44184-024-00056-z>.
232. Yang, T.; Xu, W.; Wang, L.; Petrovna, M.O. A deep learning model for psychological support in student entrepreneurship. *IEEE Access* **2025**, *13*, 40931–40942. <https://doi.org/10.1109/ACCESS.2025.3547833>.
233. Wang, F.; Dong, J. Application of an improved AFCNN model for children's psychological emotion recognition. *Sci. Rep.* **2025**, *15*, 24138. <https://doi.org/10.1038/s41598-025-10269-7>.
234. Lu, Q.; Wang, M.; Zuo, Y.; Tang, Y.; Zhang, R.; Zhang, J. Construction and verification of a risk-prediction model of psychological distress in psychiatric nurses. *BMC Nurs.* **2025**, *24*, 161. <https://doi.org/10.1186/s12912-025-02796-5>.
235. Yue, P.; Zhang, J. Learning engagement and psychological flexibility among Chinese adolescents: A moderated mediation model. *Front. Psychol.* **2025**, *15*, 1407707. <https://doi.org/10.3389/fpsyg.2025.1407707>.
236. Xu, J.; Xu, Y. The impact of self-efficacy on psychological resilience in EFL learners: A serial mediation model. *BMC Psychol.* **2025**, *13*, 858. <https://doi.org/10.1186/s40359-025-03236-4>.
237. Kamarudin, N.Y.; Kosnin, A.M.; Handayani, L.; Mamat, C.R. Structural equation modelling: Validation of a career-readiness model using psychological constructs. *Int. J. Eval. Res. Educ.* **2024**, *13*, 159–169. <https://doi.org/10.11591/ijere.v13i1.27143>.
238. Xu, C.; Wu, C.F.; Xu, D.D.; Lu, W.Q.; Wang, K.Y. Challenges to student interdisciplinary learning effectiveness: An empirical case study. *J. Intell.* **2022**, *10*, 88. <https://doi.org/10.3390/jintelligence10040088>.
239. Santana-Perera, B.; García-Barceló, C.; González Arcas, M.; Gil, D. Exploring predictive insights on student success using explainable machine learning. *Information* **2025**, *16*, 763. <https://doi.org/10.3390/info16090763>.
240. Tao, P.; Xu, R.; Song, W.; Qu, G.; Wu, H. DPPA: A deep-learning-based physiological and psychological assessment model for firefighter training injury. *IEEE Internet Things J.* **2025**, *12*, 34599–34602.
241. Prykhodko, I.; Matsehora, Y.; Kolesnichenko, O.; Stasiuk, V.; Bolshakova, A.; Bilyk, O. Psychological first aid for military personnel in combat operations: The Ukrainian model. *Mil. Behav. Health* **2021**, *9*, 289–296. <https://doi.org/10.1080/21635781.2020.1864530>.
242. Morgan, E.M.K. Including the soldier in military simulations: Modeling the psychological effects of anti-personnel landmines. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, St. Paul, MN, USA, 8–12 October 2001; pp. 1743–1747. <https://doi.org/10.1177/154193120104502409>.
243. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 4765–4774.
244. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
245. Matz, S.C.; Kosinski, M.; Nave, G.; Stillwell, D.J. Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12714–12719. <https://doi.org/10.1073/pnas.1710966114>.
246. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. <https://doi.org/10.1038/s42256-019-0114-4>.

247. Blasi, D.E.; Henrich, J.; Adamou, E.; Kemmerer, D.; Majid, A. Over-reliance on English hinders cognitive science. *Trends Cogn. Sci.* **2022**, *26*, 1153–1170. <https://doi.org/10.1016/j.tics.2022.09.015>.
248. Haotian, T.; Zeqing, L.; Zhen, W. Who is more trustworthy, AI or humans? Modeling human–AI interactions in repeated trust games with reinforcement learning. *J. Psychol. Sci.* **2025**, *48*, 920–932. <https://doi.org/10.16719/j.cnki.1671-6981.20250413>.
249. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.

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.