frontiers in Artificial Intelligence

TYPE Hypothesis and Theory
PUBLISHED 28 October 2026
DOI 10.3389/frai.2026.1899109

Check for updates

OPEN ACCESS

EDITED BY
Francesco Giannini,
Scuola Normale Superiore Classe di Scienze, Italy

REVIEWED BY
Wen-Ran Zhang,
Independent Researcher, Statesboro, United States
Vaibhavi Tiwari,
Montclair State University, United States

*CORRESPONDENCE
Jianguo Ding
jianguo.ding@bth.se
Huansheng Ning
ninghuansheng@ustb.edu.cn

RECEIVED 03 June 2026
REVISED 09 October 2026
ACCEPTED 12 October 2026
PUBLISHED 28 October 2026

CITATION
Chen R, Zhou L, Ding J and Ning H (2026) Cyberlogic-enabled Thinking Space: a closed-loop human-cyber cognitive architecture in CPST. *Front. Artif. Intell.* 9:1899109. doi: 10.3389/frai.2026.1899109

COPYRIGHT
© 2026 Chen, Zhou, Ding and Ning. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cyberlogic-enabled Thinking Space: a closed-loop human-cyber cognitive architecture in CPST

Ruizi Chen[1], Lei Zhou[1], Jianguo Ding[2]* and Huansheng Ning[1]*

[1]School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing, China, [2]Department of Computer Science, Blekinge Institute of Technology, Karlskrona, Sweden

Human–AI collaboration in Cyber–Physical–Social–Thinking (CPST) environments requires coordination among perception, reasoning, human judgment, and governance. This paper proposes Cyberlogic-enabled Thinking Space (CyTS), a conceptual governance-centric architecture organized through Entity, Mapping, Reasoning, and Governance layers. Its closed loop connects multimodal representation, heterogeneous reasoning, collaborative decisions, and controlled adaptation. The Entity layer provides operational descriptions rather than a sufficient ontology of cognition; logical consistency and predictive accuracy do not by themselves establish the causal validity of an action. CyTS therefore distinguishes associative, deductive, interventional, and exploratory claims and requires explicit assumptions, evidence provenance, uncertainty, and verification status at component interfaces. Structural causal models provide one implementable route, while equilibrium-based causal logic and proposals concerning brain–universe similarity and entangled machine thinking are examined as research perspectives with separate validation requirements. Governance mechanisms support privacy, accountability, human oversight, and runtime monitoring, including abstention when evidence or constraints are insufficient. Research hypotheses and evaluation protocols address causal transparency, decision quality, adaptation, cross-space coordination, and energy cost. Healthcare, education, and transportation scenarios illustrate the proposed workflow; they do not constitute empirical validation. CyTS contributes an auditable organizational framework and a testable implementation agenda, without claiming a completed theory of cognition or demonstrated performance gains.

KEYWORDS
closed-loop cognition, cognitive augmentation, cyberlogic, Cyberlogic-enabled Thinking Space (CyTS), Cyber-Physical-Social-Thinking (CPST), human–cyber collaboration

1 Introduction

Advances in artificial intelligence, multimodal learning (Roy et al., 2019), brain–computer interfaces (Lorach et al., 2023; Wilson et al., 2025), and ubiquitous sensing are transforming human–machine interaction from tool-based automation toward cognitive collaboration and joint decision-making (Flesher et al., 2021; Wang et al., 2025). Large foundation models provide strong semantic understanding (Vaswani et al., 2017), knowledge organization, and reasoning capabilities, while behavioral and physiological sensing enables partial modeling of human cognitive states. Consequently, AI is evolving

from an execution tool into an active participant in perception, reasoning, and decision support.

Despite these advances, reliable integration of reasoning, feedback, and accountability remains a practical challenge (Vaccaro et al., 2024). Human participation and interpretability (Reverberi et al., 2022; Yang et al., 2023) can be incorporated at multiple stages, but their presence alone does not establish that a decision is causally justified or that a physical action is safe. CPST systems therefore need explicit connections among evidence, model assumptions, decision authority, and observed outcomes, particularly when representations and actions cross domain boundaries.

These limitations become more pronounced in Cyber–Physical–Social–Thinking (CPST) environments (Ning and Liu, 2015), where cognition emerges from interactions among physical, social, cyber, and thinking spaces. Although cyber–physical and cyber–social systems have been extensively studied, cyber-enabled thinking space still lacks a systematic organizational architecture for trustworthy human–cyber cognition.

To address this gap, this paper introduces Cyberlogic as a structured organizational framework integrating perception, representation, reasoning, governance, and feedback into a unified cognitive process. Based on Cyberlogic, we further propose Cyberlogic-enabled Thinking Space (CyTS), a governance-centric human–cyber cognitive architecture for CPST environments. CyTS supports cognitive augmentation by coordinating human cognition, cyber intelligence, and governance within a closed-loop process while preserving human agency, interpretability, accountability, and adaptive feedback.

The central research question is how a governance architecture can connect heterogeneous reasoning to accountable action without confusing a coherent representation with a validated causal explanation. Two subsidiary questions concern how components should expose their assumptions and how improvements should be tested against their computational cost. We use "causal-logical transparency" to mean an inspectable connection between a claim, its assumptions, inference steps, and intervention-relevant evidence, rather than the fluency of a generated explanation. CyTS is proposed as an architectural specification; the present study does not report an implemented system or controlled experiments.

The main contributions of this paper are summarized as follows:

1. A governance-centric cognitive architecture, CyTS, is proposed for CPST environments, integrating perception, representation, reasoning, human participation, and feedback into a unified closed-loop process;
2. A four-layer Entity-Mapping-Reasoning-Governance structure is introduced to operationalize Cyberlogic and support coordination between human cognition and cyber intelligence;
3. Governance is embedded directly into the reasoning process through privacy protection, accountability management, human oversight, responsibility allocation, and runtime monitoring mechanisms;
4. Research hypotheses, validation criteria, and operational workflows are provided to improve the testability and implementability of the proposed framework;
5. Cross-domain mappings in healthcare, education, and transportation demonstrate the architectural consistency and potential applicability of CyTS.

Section 2 reviews related architectures and the epistemic limits of their reasoning mechanisms. Section 3 defines CyTS and its hypotheses. Sections 4 and 5 specify the architecture and governance mechanisms. Section 6 presents operational workflows and domain illustrations. Section 7 sets out evaluation requirements and limitations, and Section 8 concludes the paper.

2 Related work and structural limitations

2.1 Cognitive architectures, distributed cognition, and hybrid intelligence

Cognitive architectures provide models of learning, reasoning, and computational cognition, with particular emphasis on the organization of individual agents (Ning, 2016; Zeng et al., 2020). Distributed cognition broadens the unit of analysis to interactions among people, tools, artifacts, and environments, while hybrid intelligence emphasizes complementary human and machine capabilities (Glickman et al., 2025; Ning et al., 2017). Together, these perspectives motivate the integration of human judgment and cyber computation in CPST environments.

For CyTS, the relevant challenge is to translate these complementary perspectives into explicit cross-space interfaces. An account of cognitive organization or task complementarity does not by itself specify how uncertain observations, reasoning assumptions, human authority, and feedback should be connected in an executable workflow. CyTS builds on these perspectives by proposing an organizational structure for this coordination; it does not claim to replace established cognitive theories or to establish a complete theory of human thinking.

2.2 Human–AI governance and cross-space integration

Human-centered AI emphasizes augmentation, reliability, transparency, and human control. Human-in-the-Loop AI, Human–AI Teaming, Interactive Machine Learning, and Mixed-Initiative Systems operationalize these objectives through feedback, task coordination, and shared decision-making. Work on human-like learning and reasoning provides a complementary cognitive perspective (Lake et al., 2017). Causal reasoning (Pearl, 2018), cyberentity modeling (Dhelim et al., 2020), and explainable AI (Pahud de Mortanges et al., 2024) address related requirements for representing and interpreting decisions. Non-invasive BCIs can supply additional cognitive inputs (Edelman et al., 2025), while bounded rationality and free-energy accounts inform models of cognition and adaptation (Kahneman, 2003; Friston, 2010); none of these contributions alone specifies an end-to-end governance architecture.

Governance-oriented approaches address different parts of this integration problem. Constitutional AI uses principles to regulate model behavior; trustworthy AI frameworks articulate lifecycle requirements; and runtime assurance monitors operation and supports intervention within specified conditions (Schierman et al., 2020). These approaches overlap, and particular implementations may already embed constraints and feedback within inference. Governance integration alone therefore does not establish CyTS's novelty or superiority. Its proposed contribution is the explicit connection among CPST representations, reasoning evidence, human authority, and controlled adaptation. Table 1 compares architectural emphases rather than experimentally measured capabilities; the CyTS column states intended functions.

This comparison yields four integration requirements: governance must be connected to intermediate inference, feedback must preserve decision provenance, representations must retain temporal and contextual information, and cross-space actions must remain subject to authority and safety constraints. These are implementation-dependent requirements, not universal deficiencies of existing architectures. Additional monitoring and verification may be justified for consequential decisions but can impose unnecessary overhead in routine, low-risk tasks. Beyond organizational integration, CyTS must also distinguish logically valid statements from empirically supported intervention claims. The following subsections examine this epistemic boundary, alternative causal formalisms, and exploratory perspectives on machine cognition.

TABLE 1 Architectural emphases and the proposed contribution of CyTS.

Approach	Primary emphasis	Relationship to CyTS
Human-centered/human-in-the-Loop AI	Human control and feedback across design and operation	Supplies oversight methods; CyTS specifies cross-space evidence and decision interfaces.
Human–AI teaming/hybrid Intelligence	Complementary capabilities, coordination, and task allocation	Supplies collaboration methods; CyTS links allocation to governance records.
Constitutional AI	Principles governing model behavior	Supplies policy constraints; CyTS additionally coordinates physical evidence and action review.
Runtime assurance	Monitoring and intervention within a defined operating envelope	Supplies safety mechanisms; CyTS links them to reasoning assumptions and human escalation.
Cognitive digital twins	Representation and state synchronization	Supplies state models; CyTS additionally specifies evidence status and decision authority.
CyTS	CPST-wide coordination of representation, reasoning, governance, and feedback	Intended integration; comparative benefits require implementation and controlled evaluation.

2.3 Logical validity, causal evidence, and computational cost

An entity-centered description identifies objects and attributes, while truth-valued inference establishes consequences relative to premises. Neither alone supplies a model of how an intervention changes a physical or social process. This limitation does not imply that symbolic methods are intrinsically unsuitable for AI: explicit rules remain useful for checking consistency, permissions, and constraints. The missing step is the connection between formal symbols, measured processes, and action-dependent outcomes. Likewise, a probabilistic predictor estimates regularities under its data-generating conditions; predictive success need not transfer when an action changes those conditions.

LLMs can assist with causal hypotheses and formalization, but generated reasoning is not automatically a faithful account of the computation that produced an answer. Experiments demonstrate misleading chain-of-thought explanations (Turpin et al., 2023), and CLadder separates associational, interventional, and counterfactual questions to test formal causal reasoning (Jin et al., 2023). These results motivate independent checks of intermediate claims, rather than a blanket assertion that LLMs cannot reason. In CyTS, an explanation must distinguish retrieved evidence, model assumptions, deductive consequences, and unverified suggestions.

Causal opacity also complicates the diagnosis of unnecessary generation, repeated calls, and failed action proposals. However, opacity is not an established universal cause of high energy consumption. Deployment measurements show that energy depends on task and model configuration (Luccioni et al., 2024); adding verification can itself increase cost. We therefore treat any efficiency benefit of transparent reasoning as a hypothesis to test at matched decision quality and safety.

2.4 Causality beyond truth-valued representation

Pearl's (2018) structural causal framework distinguishes observation from intervention and counterfactual inference. It is not merely a restatement of empirical correlation or Boolean truth. Its conclusions are conditional on structural assumptions, intervention semantics, and identification conditions; data alone do not generally determine those assumptions. Thus, CyTS must expose possible confounding, model misspecification, and the limits of transporting a causal claim across environments.

An alternative line of work seeks logically definable causality through equilibrium-based regularity. Zhang and Peace (2014) propose bipolar dynamic logic (BDL) and bipolar quantum linear algebra (BQLA), framing causation through bipolar interactions rather than only truth-valued descriptions of beings. Their account is positioned against the limitations they identify in Aristotelian and Humean treatments of causation. Here, "formal causality" denotes derivability within a specified formal system; it is not used as a

synonym for Aristotle's formal cause. This perspective broadens CyTS's candidate reasoning formalisms, but a derivation in BDL does not by itself establish that an encoded CPST mechanism is empirically correct.

Equilibrium also appears within conventional dynamical causal modeling. Mooij et al. (2013) establish conditions under which equilibria of differential equations admit structural causal representations. This connection should not be conflated with BDL's broader foundational claims. In either approach, CyTS requires a specified state space, interaction rules, intervention semantics, and conditions for existence, uniqueness, and stability where relevant. An observed balance or a fitted fixed point alone is insufficient evidence of a causal mechanism.

2.5 Brain–universe similarity and exploratory quantum perspectives

Vazza and Feletti (2020) report quantitative similarities between selected neuronal-network and cosmic-web structural measures, using proximity-based approximations of neural connectivity. Such observations can motivate hypotheses about network organization; they do not establish shared cognitive functions or a common causal mechanism. CyTS therefore treats structural analogy as a source of candidate representations, not as evidence that a cosmic or quantum model explains human thinking.

Zhang (2024) proposes global realism with bipolar strings, a causal-logical account of quantum gravity, and causal-logical brain modeling for entangled machine thinking and imagination (EMTI), including a proposed mind–light–matter unity. These proposals connect foundational physics, quantum information science, and machine cognition. For CyTS, they motivate exploratory relational reasoning modules. Their adoption would require an executable specification and discriminating tests beyond the architectural illustrations presented here. Computational coupling of representations must also be distinguished from experimentally demonstrated physical quantum entanglement. CyTS neither assumes a quantum substrate for cognition nor claims to validate quantum gravity.

3 Cyberlogic-oriented structural foundation of CyTS

3.1 Definition and organizational role of Cyberlogic

Cyberlogic is a foundational framework proposed for organizing cross-space reasoning, logical coordination, and governance within the Cyber-Physical-Social-Thinking (CPST) hyperspace (Ning et al., 2023, 2017). Unlike conventional AI reasoning frameworks that primarily focus on inference efficiency or model performance, Cyberlogic emphasizes the structural organization, traceability, semantic consistency, and governance of reasoning processes across heterogeneous spaces (Durstewitz et al., 2019). Within the Cyberlogic framework, logical activities are organized through four interconnected layers:

- **Entity layer:** the Entity layer defines the cyberentities participating in reasoning processes, including physical entities, social entities, cognitive entities, and cyber-enabled entities. Cyberlogic treats entities as structured logical carriers with unified attribute schemas supporting identity, spatiotemporal states, social semantics, cognitive states, and security constraints.
- **Mapping layer:** the Mapping layer transforms entities, relationships, and cognitive states from physical, social, and thinking spaces into cyber representations capable of participating in logical operations. Cyberlogic explicitly emphasizes that mapping is not equivalent replication, but selective reconstruction with uncertainty and semantic abstraction.
- **Reasoning layer:** the Reasoning layer coordinates multiple logic sources during inference, including both cyberized logic migrated from PST spaces and endogenous cyberlogic generated within cyberspace through data-driven and generative mechanisms. Rather than unifying all logic into a single reasoning formalism, Cyberlogic focuses on multi-logic coordination, reasoning traceability, uncertainty propagation, and conflict resolution.
- **Governance layer:** the Governance layer provides meta-level regulation for the entire reasoning process. It ensures semantic consistency, traceability, integrity maintenance, responsibility allocation, and controllable evolution of reasoning systems through governance constraints and operational principles (Mittelstadt et al., 2016).

The Entity layer is an engineering device for identifying participants, observations, and responsibilities; it does not commit CyTS to a being-qua-being account in which an ontology of entities suffices to explain cognition. Processes, events, interventions, and time-dependent relations must be represented alongside entities. Similarly, "multi-logic" denotes coordination among explicitly specified methods, not a new universal calculus. A method may establish formal validity within its assumptions while leaving empirical adequacy unresolved. Governance must retain that distinction when authorizing actions.

Within this theoretical foundation, CyTS can be understood as a Cyberlogic-oriented cognitive architecture specifically focused on governance-aware augmentation of the Thinking Space in CPST environments. Rather than replacing human cognition, CyTS extends Cyberlogic toward recursive human–cyber collaborative cognition, adaptive reasoning coordination, and cognitive governance under multi-space interaction conditions. To clarify the theoretical position of CyTS within CPST, Figure 1 illustrates how Cyberlogic functions as the organizational logic that connects physical inputs, social contexts, cyber models, and human cognition into a closed-loop cognitive structure.

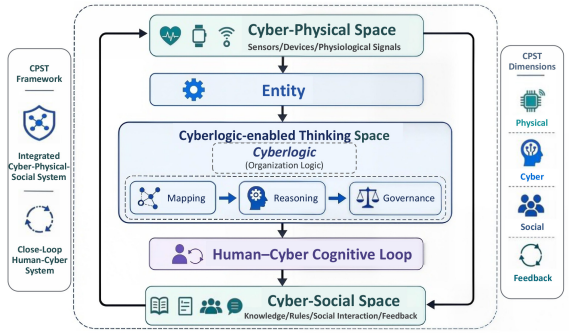


FIGURE 1
Cyberlogic-driven CyTS within the CPST framework. Cyberlogic provides the organizational logic that structures and orchestrates the cyberlogic-enabled thinking space (CyTS), linking physical inputs, cyber reasoning, and social feedback into a closed-loop human–cyber cognitive system.

3.2 Definition of Cyberlogic-enabled Thinking Space

Equation 1 defines CyTS through the following architectural tuple:

$$CyTS = \langle E, M, R, G, H, C \rangle. \tag{1}$$

This notation records components and their intended interactions; it does not constitute an algebra with proven closure or convergence properties. Equations 2–5 provide a minimal state-transition specification:

$$S_t = M(E_t), \tag{2}$$

$$D_t = R(S_t, K, U_t; \Pi_t), \tag{3}$$

$$D'_t = G(D_t, HC_t \mid E_x, R_c, H_j, \Pi_t), \tag{4}$$

$$S_{t+1} = \mathcal{F}(S_t, D'_t, F_t). \tag{5}$$

Here, E_t comprises timestamped inputs; S_t is a structured state whose numerical encoding is implementation-dependent; K contains domain knowledge; and U_t records uncertainty. The policy specification Π_t constrains inference as well as action selection, ensuring that governance is not confined to an output filter. D_t contains candidate decisions and their evidence records, HC_t records human–cyber collaboration, and D'_t is an authorized decision or an abstention/escalation outcome. E_x, R_c, and H_j denote explainability, risk-control, and human-judgment constraints. Environmental feedback F_t supports a controlled transition $\mathcal{F}$; safety and stability require separate verification.

This definition emphasizes three points. First, human thinking remains the cognitive subject. CyTS does not replace human cognition, but provides cyber-enabled cognitive support. Second, cyber cognitive models serve as augmentation components. Their role is to expand representation, reasoning, and decision-support capabilities, rather than to become autonomous cognitive agents. Third, human-cyber collaboration mechanisms provide the structural connection between human cognition and cyber cognition (Pariser, 2011). They enable bidirectional interaction, feedback updating, and governance regulation. Thus, CyTS is not equivalent to artificial consciousness or strong artificial intelligence. It is a cognitive augmentation structure that combines human cognitive agency with cyber computational capability under governance constraints. To improve formal consistency and engineering interpretability, the core mathematical symbols and functional variables used throughout the CyTS framework are summarized in Table 2.

3.3 Cyberlogic-to-CyTS structural mapping

Cyberlogic provides the structural basis for implementing CyTS. The four Cyberlogic layers can be mapped into four functional layers of CyTS. Based on this theoretical position, the four Cyberlogic layers can be further mapped into the functional architecture of CyTS, as shown in Figure 2. The detailed mapping relationship between Cyberlogic and CyTS is summarized in Table 3.

This mapping transforms Cyberlogic from a theoretical framework into an implementable cognitive architecture. Unlike conventional AI pipelines, the Governance layer is not placed outside the system. Instead, it is embedded as a structural component that regulates reasoning outputs, feedback updating, and human–cyber collaboration (Rahwan et al., 2019; Dellermann et al., 2019).

3.4 CyTS within the CPST coupling structure

CyTS does not operate independently. It is dynamically coupled with cyber-enabled physical space and cyber-enabled social space.

Cyber-enabled Physical Space provides objective data from sensors, devices, environments, and biological signals. Cyber-enabled Social Space provides social knowledge, interaction records, collective experience, institutional rules, and domain norms (Amershi et al., 2019). CyTS integrates these inputs through cyber cognitive models and generates decision support for human cognition.

The CPST coupling process can be summarized as follows. Physical and social signals are first collected through the Entity layer. These signals are transformed into structured cognitive representations through the Mapping layer. The reasoning layer performs inference and generates recommendations. The Governance layer then regulates the output through human review, explainability, ethical constraints, and feedback mechanisms. The final decision influences physical actions and social interactions, thereby re-entering the CPST loop (Brundage et al., 2020; Ning et al., 2018).

3.5 Research hypotheses and validation criteria

The following hypotheses concern future implementations, not results of the present conceptual study. Comparisons should hold the underlying model, data access, action set, and task conditions

TABLE 2 Mathematical symbols and functional definitions in CyTS.

Symbol	Definition	Functional meaning
$CyTS$	Cyberlogic-enabled Thinking Space	Governance-aware human–cyber cognitive architecture
E	Entity perception space	Multimodal cognitive input acquisition
M	Cognitive mapping function	Transformation from raw signals to structured representations
R	Multi-logic reasoning function	Hybrid reasoning and inference coordination
G	Governance feedback function	Governance-aware cognitive regulation
H	Human cognitive participation	Human judgment, supervision, and feedback
C	Cyber cognitive models	AI-based and cyber computational reasoning models
S	Structured cognitive state	Internal semantic representation at time t
S_{t+1}	Updated cognitive system state	Adaptive cognitive state after feedback updating
K	Knowledge constraints	Knowledge graph and domain-specific constraints
U	Uncertainty states	Probabilistic uncertainty representation
D	Reasoning outputs	Intermediate inference and recommendation results
D'	Governance-regulated outputs	Final regulated decisions
E_t	Multimodal entity inputs at time t	Dynamic perception inputs
HC	Human-cyber collaborative cognition	Hybrid human-AI cognitive state
F_t	Environmental feedback	External adaptive feedback signals
$\alpha, \beta, \gamma, \delta$	Reasoning weights	Adaptive contribution coefficients
λ	Human-cyber collaboration coefficient	Relative cognitive authority allocation
ϕ_e	Explainability governance weight	Governance emphasis on interpretability
ϕ_r	Risk-control governance weight	Governance emphasis on safety
ϕ_h	Human-judgment governance weight	Governance emphasis on human oversight
E_x	Explainability constraints	Constraints for transparent reasoning
R_c	Risk-control constraints	Safety and ethical regulation constraints
H_j	Human judgment constraints	Human supervisory and intervention constraints

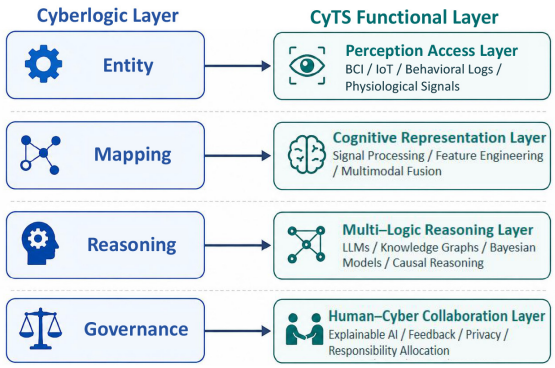


FIGURE 2
Cyberlogic-to-CyTS four-layer architecture. The cyberlogic layers of entity, mapping, reasoning, and governance are systematically mapped onto the perception access, cognitive representation, multi-logic reasoning, and human–cyber collaboration layers of CyTS, respectively, forming a governance-aware closed-loop human–cyber cognitive architecture.

constant where possible. All thresholds, primary outcomes, and minimally meaningful effects must be specified before evaluation.

H1 (causal-logical transparency): Compared with the same decision-support model using *post-hoc* explanations, CyTS increases the proportion of decisions with independently checkable evidence and assumption records. The traceability rate is defined in Equation 6 as follows:

$$TR = N_{\text{traceable}}/N_{\text{total}}, \tag{6}$$

Record completeness must be assessed separately from correctness. Primary checks include validity of formal steps, explicit identification assumptions for intervention queries, and faithful correspondence between recorded steps and the executed workflow. SHAP values and perceived explanation quality are supplementary indicators, not causal-validity measures.

H2 (decision quality): CyTS improves task utility relative to both human-only and AI-only baselines under predefined uncertainty conditions, without exceeding a safety non-inferiority margin. Measure task success, severity-weighted errors, abstention, coverage, and error recovery; include the human time needed to review evidence.

H3 (adaptation): governance-controlled updating improves recovery after a defined distribution shift relative to a fixed-model baseline, without unacceptable oscillation, constraint violations, or rollback frequency. Specify the disturbance, observation horizon, and acceptable recovery region in advance. No convergence guarantee follows from the architecture alone.

H4 (cross-space coordination): explicit temporal and semantic contracts reduce inconsistent cross-space decisions while keeping end-to-end latency within the application's deadline. Measure unit and timestamp mismatches, synchronization failures, and deadline violations. Additional checks may increase latency, so consistency and timeliness are evaluated jointly.

H5 (resource efficiency): relative to an otherwise identical unconstrained generation workflow, selective reasoning and verification reduce energy per accepted correct decision at matched

TABLE 3 Cyberlogic-to-CyTS structural mapping.

Cyberlogic layer	CyTS functional layer	Core function	Representative technologies
Entity	Perception access layer	Acquisition and identification of cognition-related entities and signals	BCI, behavioral logs, IoT sensors, wearable devices
Mapping	Cognitive representation layer	Transformation of raw cognitive data into structured features and representations	Signal processing, feature engineering, multimodal fusion, embeddings
Reasoning	Multi-logic reasoning layer	Inference, prediction, explanation, and decision support	Large models, knowledge graphs, Bayesian models, causal reasoning
Governance	Human–cyber collaboration layer	Human feedback, responsibility allocation, ethical control, and adaptive updating	Explainable AI, human-in-the-loop interfaces, privacy protection, access control

safety and coverage. Measure total pipeline energy, including retrieval, inference, verification, monitoring, and retries, alongside latency and human review time. A shorter explanation or fewer output tokens is not sufficient evidence of an energy saving.

Effect sizes and confidence intervals should be reported with prespecified acceptance margins, sampling units, and multiplicity controls. A nonsignificant superiority test is inconclusive rather than proof of equivalence or falsification. A claimed benefit is unsupported when its prespecified criterion is not met and contradicted when sufficiently precise estimates exclude the minimally meaningful benefit or reveal an unacceptable trade-off. Observational datasets such as MIMIC-IV may support retrospective workflow assessment, but cannot alone establish the effects of deploying an intervention. Causal tests require identified designs or simulations with known intervention ground truth.

4 Cyberlogic-driven cognitive architecture of CyTS

4.1 Overall architecture

The implementation of CyTS is not the deployment of a single AI algorithm. It is a Cyberlogic-driven cognitive architecture composed of perception access, cognitive representation, multi-logic reasoning, and human–cyber collaboration layers.

The architecture follows the principle that perception, reasoning, and governance should not be separated (Nguyen et al., 2014). Instead, they should be integrated into a closed-loop cognitive system. The system receives cognitive signals from humans, physical environments, and social contexts; transforms them into structured representations; conducts multi-logic reasoning; generates interpretable recommendations; receives human feedback; and updates models or constraints accordingly (Floridi et al., 2018).

Although CyTS is organized through four fundamental layers, the architecture is designed to support future scalability and extension. Adaptive evolution is considered a cross-layer capability rather than an independent Adaptation Layer. Through closed-loop feedback among Entity, Mapping, Reasoning, and Governance layers, CyTS can continuously adjust cognitive representations, reasoning strategies, and collaboration mechanisms according to environmental changes and human feedback. This design preserves the conceptual integrity of the Cyberlogic framework while enabling flexible adaptation in evolving CPST scenarios.

Furthermore, future implementations may introduce additional cross-layer interaction pathways for specific application requirements. For example, direct verification channels between the Entity and Governance layers can support privacy protection, security monitoring, and runtime compliance checking. Such extensions enhance architectural scalability without altering the fundamental Entity–Mapping–Reasoning–Governance organization.

4.2 Perception access layer

The perception access layer corresponds to the Entity layer of Cyberlogic. Its function is to identify and acquire cognition-related entities and signals. These signals may include brain signals, physiological indicators, behavioral logs, language inputs, environmental data, and social interaction records.

Brain-computer interfaces provide a direct or indirect channel between human cognitive states and cyber systems (Patrick-Krueger et al., 2025; Edelman et al., 2019). Non-invasive EEG-based BCIs can capture attention, fatigue, workload, or simple intention-related signals. However, current BCI technologies remain limited in spatial resolution, robustness, and semantic decoding capability. Therefore, within CyTS, BCI should be understood as an auxiliary cognitive input channel rather than a mind-reading technology (Lotte et al., 2013).

Behavioral data are also important cognitive inputs. Interaction logs, click sequences, writing behavior, browsing paths, response time, and task-switching patterns can reflect cognitive rhythm, decision strategies, and workload. Compared with neural signals, behavioral data are easier to collect and more suitable for long-term cognitive modeling.

4.3 Cognitive representation layer

The cognitive representation layer corresponds to the Mapping layer of Cyberlogic. Its function is to transform heterogeneous raw signals into structured cognitive representations (Moor et al., 2023; Huang et al., 2023).

Different data sources require different mapping strategies. EEG signals can be transformed into frequency-domain, time-frequency, and connectivity features. Behavioral logs can be transformed into temporal patterns, sequence features, and interaction profiles. Language inputs can be transformed into semantic embeddings, contextual vectors, and intent representations. Multimodal fusion can integrate these features into unified cognitive representations (Sundrani et al., 2023; Shi et al., 2025).

Each mapped object should carry a source identifier, timestamp, units, transformation history, and uncertainty descriptor. Relations must state whether they encode observed association, a domain rule, or an assumed causal mechanism. This prevents an embedding similarity or knowledge-graph edge from silently becoming an intervention claim. Mapping failures, incompatible units, and stale observations are explicit reasons to request new evidence or abstain.

4.4 Multi-logic reasoning layer

The multi-logic reasoning layer corresponds to the reasoning layer of cyberlogic. Its function is to integrate multiple reasoning paradigms to generate inference results, explanations, and decision recommendations. To conceptually illustrate the composition of the reasoning process, the reasoning outcome can be represented as:

$$D_t = \alpha \cdot L_{\text{data}}(S_t) + \beta \cdot L_{\text{knowledge}}(S_t, K) + \gamma \cdot L_{\text{uncertainty}}(S_t, U_t) + \delta \cdot L_{\text{human}}(S_t, H) \tag{7}$$

$$\alpha + \beta + \gamma + \delta = 1 \tag{8}$$

The reasoning coefficients may be adapted to runtime context and normalized as aggregation weights. Entropy of an explicitly specified uncertainty distribution and semantic compatibility with domain knowledge are possible routing signals, but their suitability and scaling require empirical calibration. Normalization alone does not give these weights a causal or probabilistic interpretation. One possible normalization is:

$$w_i = \frac{\exp(\tilde{w}_i)}{\sum_{j \in \{\alpha,\beta,\gamma,\delta\}} \exp(\tilde{w}_j)}, \qquad i \in \{\alpha, \beta, \gamma, \delta\} \tag{9}$$

Here, $\tilde{w}_i$ is a finite context-dependent score; normalization gives positive weights summing to one. It does not guarantee decision quality.

In practical implementations, the large models component within the reasoning layer is not limited to a specific model type. It may include general-purpose foundation models, such as large language models, as well as domain-adapted models customized for specific CPST applications. General-purpose models provide broad semantic understanding and reasoning capabilities, while domain-specific models, knowledge graphs, and specialized inference modules can be integrated to improve contextual accuracy and reliability in application scenarios such as healthcare and education. Therefore, the reasoning layer should be regarded as a heterogeneous reasoning framework that coordinates multiple models and knowledge sources rather than a direct deployment of a single large model.

Uncertainty logic supports probabilistic updating under incomplete information, while data-driven methods provide representation learning and predictive inference (Mnih et al., 2015; Geirhos et al., 2020). Knowledge-structured methods expose semantic relations and domain constraints; causal use additionally requires an explicit mechanism and intervention semantics. Human collaboration contributes contextual judgment and values. CyTS coordinates these roles while preserving the distinction between predictive, deductive, causal, and normative claims (Bender et al., 2021).

The weighted expression in Equations 7–9 is meaningful only when component outputs are mapped to comparable action scores with a declared calibration procedure. It must not add proofs, probabilities, and human judgments as though they were interchangeable quantities. The weights are routing or aggregation parameters, not evidence that the combined output is causally valid. Hard constraints and verified contradictions take precedence over weighted scores. In particular, knowledge-graph structure supplies causal information only when its edges have explicit causal semantics and supporting assumptions.

For every candidate, the Reasoning layer supplies an evidence record

$$\mathcal{A}_t = \langle q, \mathcal{T}, \mathcal{M}, \mathcal{B}, \mathcal{P}, \mathcal{V}, U_t \rangle,$$

where q is the query, $\mathcal{T}$ its claim type (associative, deductive, interventional, counterfactual, or exploratory), $\mathcal{M}$ the model and version, $\mathcal{B}$ the assumptions, $\mathcal{P}$ the evidence and derivation provenance, and $\mathcal{V}$ the verification status. An LLM may propose a causal graph or translate a question, but an intervention claim must additionally identify its estimand and assumptions and pass an appropriate solver or domain check. A BDL/BQLA module would use the same contract while declaring its bipolar encoding and inference rules. Outputs from different formalisms remain distinguishable; unresolved disagreement triggers review rather than unqualified fusion.

4.5 Human–cyber collaboration layer

The human–cyber collaboration layer corresponds to the governance layer of cyberlogic. Its function is to regulate system outputs, support human decision-making, collect feedback, and maintain responsibility boundaries. The human-cyber collaborative outcome can be conceptually expressed as:

$$HC = \lambda_t \cdot H + (1 - \lambda_t) \cdot C,$$
$$\lambda_t = \max(0, \min(1, \lambda_0 - \omega_1 \cdot W(E_t) + R(U_t))) \tag{10}$$

In Equation 10, H and C must denote comparable decision-support scores if numerical aggregation is used. The coefficient λ_t describes assistance allocation, not a transfer of legal or professional responsibility. Workload $W(E_t)$, risk $R(U_t)$, and the scaling parameter ω_1 require domain calibration; $R(U_t)$ here is a scalar risk index, distinct from the reasoning function R. Fatigue alone must not authorize increased autonomy. If a high-risk task exceeds the available human review capacity, the system defers, escalates to another qualified operator, or enters an independently specified fallback mode.

This mechanism enables flexible cognitive allocation between humans and cyber systems while preserving human oversight

and responsibility. Through explainable interfaces and governance mechanisms, users can review, modify, or provide feedback on system recommendations. Such feedback contributes to the refinement of cognitive representations, reasoning strategies, and governance mechanisms within the closed-loop process.

This layer ensures that cognitive augmentation does not become cognitive replacement. Human agency, responsibility, and value judgment remain structurally embedded in the system.

5 Cyberlogic mechanisms for human–cyber cognitive integration

5.1 Organizational logic of cognitive coupling

The central function of Cyberlogic in CyTS is to organize cognitive coupling. Conventional AI systems often separate data acquisition, model inference, user interaction, and governance regulation. This separation can complicate feedback coordination, responsibility allocation, and adaptation.

Cyberlogic overcomes this limitation by integrating the four layers into one cognitive loop. An entity enables cognitive signals to enter the system. Mapping transforms them into structured representations. Reasoning generates inference through multi-logic coordination. Governance regulates outputs and introduces human feedback into the system. CyTS's Cyberlogic-driven operation is supported by multiple complementary logic mechanisms. As summarized in Table 4, these mechanisms are not independent modules, but are organized across the mapping, reasoning, and governance layers. Based on this multi-logic structure, the following subsections further analyze the uncertainty, data-driven, and human–cyber collaborative logics in CyTS.

5.2 Uncertainty logic

Cognitive tasks often involve incomplete information, dynamic environments, and ambiguous evidence. Therefore, CyTS requires uncertainty logic to support probabilistic reasoning (Gao et al., 2021). In medical diagnosis, risk assessment, and dynamic decision-making, Bayesian inference and uncertainty quantification can support continuous belief updating in response to new evidence.

Within CPST, evidence may come from physical, social, and thinking spaces. Therefore, uncertainty logic enables CyTS to integrate heterogeneous information and continuously update beliefs in the presence of incomplete information.

5.3 Data-driven logic

Data-driven logic supports representation learning, prediction, and generation. Neural networks, large language models, and multimodal models can learn complex mappings from high-dimensional data and generate useful cognitive representations (Willett et al., 2023; Baltrušaitis et al., 2019).

Within CyTS, data-driven logic expands cognitive representation capability (Ngiam et al., 2011; Shrestha et al., 2022). It enables the system to identify latent patterns in behavioral logs, physiological signals, medical images, educational records, and traffic data. However, data-driven logic is primarily based on statistical correlations and may suffer from limited interpretability, hallucination, bias, and weak causal reasoning (Chang and Bergen, 2024; Gilbert et al., 2024). Therefore, data-driven logic cannot serve as the only logic of CyTS. It must be combined with uncertainty logic, knowledge-structured logic, and human–cyber collaborative logic.

5.4 Human–cyber collaborative logic

Human–cyber collaborative logic reflects the core feature of CyTS. The system does not make decisions independently, nor does it reduce humans to passive validators. Instead, humans and cyber models participate in a structured cognitive loop (Collins et al., 2024). Human feedback establishes a cycle of:

$$\text{Input} \rightarrow \text{Representation} \rightarrow \text{Reasoning} \rightarrow \text{Human Review} \rightarrow \text{Governance Feedback} \rightarrow \text{Model Update}$$

Through this loop, CyTS can continuously adapt to users, tasks, and environments while maintaining human control (Shneiderman, 2020).

5.5 Governance-centric cognitive feedback

Governance is the distinguishing feature of Cyberlogic-driven CyTS. In conventional AI systems, governance is often external to model reasoning. In CyTS, governance constraints directly regulate reasoning outputs. Governance feedback can be represented as:

$$D_t' = \arg\max_{d \in \Omega(R_c)} \Big[\phi_e \cdot f_{\text{exp}}(d \mid E_x) + \phi_h \cdot f_{\text{human}}(d \mid H_j) - \phi_r \cdot f_{\text{risk}}(d \mid R_c) \Big], \quad \sum \phi_i = 1 \tag{11}$$

The set $\Omega(R_c)$ contains actions admissible under the implemented constraints; membership is not an unconditional guarantee of real-world safety. The scores in Equation 11 require explicit scales and validation. In particular, f_{exp} should assess evidence coverage, declared assumptions, and verified reasoning status; feature attribution alone cannot certify causality. For actions requiring causal justification, the associated record $\mathcal{A}_t$ must pass the relevant identification and verification checks. If the admissible set is empty, evidence is inadequate, or a deadline is exceeded, the system abstains or invokes a domain-validated fallback and records the reason. The fallback itself must have an operating envelope and an accountable owner.

The governance formulation specifies how interpretability requirements, risk policies, and human review can regulate candidate decisions (Tonekaboni et al., 2019). Its effectiveness depends on the correctness of the implemented constraints, the quality of evidence, and the capacity of reviewers. Governance can

TABLE 4 Multi-logic mechanisms in cyberlogic-driven CyTS.

Logic mechanism	Cyberlogic position	Main function	Typical methods	Role in CyTS
Uncertainty logic	Reasoning layer	Handles incomplete information, ambiguity, and dynamic risk	Bayesian inference, probabilistic graphical models, uncertainty quantification	Supports risk-aware and adaptive cognitive inference
Data-driven logic	Mapping and reasoning layers	Learns patterns and representations from multimodal data	Neural networks, large models, multimodal learning	Expands cognitive representation and prediction capability
Knowledge-structured logic	Reasoning layer	Provides explicit semantic relations and domain constraints	Knowledge graphs, ontologies, causal models	Improves interpretability and domain consistency
Human–cyber collaborative logic	Governance layer	Integrates human judgment, feedback, and value constraints	Human-in-the-loop, explainable AI, interactive decision support	Preserves human agency and responsibility
Governance logic	Governance layer	Regulates outputs, privacy, responsibility, and ethical constraints	Privacy protection, access control, audit trails, ethical governance	Ensures trustworthy and controllable cognitive augmentation

block known violations and expose unsupported assumptions; it cannot guarantee that all harmful outcomes are anticipated.

5.6 Governance operationalization mechanisms

While governance is often discussed as a high-level principle in human–AI systems, practical deployment requires governance to be implemented through executable mechanisms that directly influence system behavior. Therefore, the Governance layer of CyTS is operationalized through five complementary governance components: Privacy Governance, Accountability Governance, Human Oversight Governance, Responsibility Allocation Governance, and Runtime Monitoring Governance.

- **Privacy governance**: privacy Governance regulates the collection, storage, access, and utilization of cognitive data throughout the CyTS lifecycle. Since CyTS may process sensitive information such as behavioral records, physiological signals, cognitive states, and decision preferences, privacy protection cannot be treated as an external compliance requirement (Kosinski, 2017). Instead, privacy constraints participate directly in data access and reasoning processes through mechanisms such as access control, anonymization, differential privacy, and secure data management.
- **Accountability governance**: accountability Governance ensures that reasoning processes remain traceable and auditable. Each recommendation generated by CyTS is associated with supporting evidence, reasoning pathways, governance interventions, and human feedback records. This traceability allows system outputs to be reviewed, explained, and audited when disputes or failures occur. Consequently, accountability is embedded within the cognitive loop rather than being evaluated only after decision execution.
- **Human oversight governance**: human Oversight Governance preserves meaningful human control over system behavior. Human participants may review, modify, reject, or override system-generated recommendations according to contextual requirements and domain regulations. Unlike conventional human-in-the-loop systems where intervention is limited to specific stages, oversight in CyTS remains available throughout the entire cognitive cycle. This mechanism prevents excessive automation and helps maintain human agency in safety-critical scenarios.
- **Responsibility allocation governance**: responsibility Allocation Governance defines the distribution of responsibilities among developers, deployers, institutions, intelligent agents, and end users. Because decisions emerge through human–machine collaboration, responsibility cannot be assigned solely to either humans or AI systems. CyTS therefore adopts traceable responsibility allocation mechanisms that clarify decision ownership and accountability across different operational roles.
- **Runtime monitoring governance**: runtime Monitoring Governance continuously evaluates system behavior during operation. It detects abnormal outputs, policy violations, unsafe recommendations, model drift, and unexpected interactions. When governance violations are identified, corrective actions may be triggered automatically or escalated to human supervisors. This mechanism enables governance to function as a continuous regulatory process rather than a static rule set.

Implementation requires concrete privacy, audit, and monitoring policies. Where differential privacy is used, the mechanism, adjacency relation, privacy budget ε_{DP}, and composition accounting must be stated; access control remains a separate requirement. Each audit record links the decision identifier and timestamp to $\mathcal{A}_t$, the governance action, human feedback, final decision, and model/policy version. Risk above a prespecified threshold triggers escalation or a validated fallback. Threshold selection, missed detections, and false alarms must be evaluated within the application operating envelope.

Collectively, these governance mechanisms transform governance from a conceptual principle into an executable component of cognition. Governance therefore participates

directly in reasoning, decision generation, feedback incorporation, and adaptive updating, forming a governance-aware cognitive loop within CyTS.

6 Closed-loop operation and representative applications

6.1 Closed-loop cognitive operation process

The Cyberlogic-driven CyTS system operates as a closed-loop cognitive process rather than a linear pipeline (Morey et al., 2025). The operational process includes multi-source cognitive input, cognitive representation, multi-logic reasoning, human–cyber collaborative decision-making, governance feedback, and adaptive updating.

CyTS first collects data from physical, social, cyber, and human cognitive contexts. These data include sensor readings, physiological signals, EEG signals, behavioral logs, language inputs, social rules, and domain knowledge. The system then transforms heterogeneous data into structured cognitive representations through signal processing, feature extraction, semantic embedding, and multimodal fusion.

The system performs reasoning by integrating uncertainty logic, data-driven logic, knowledge-structured logic, and human–cyber collaborative logic. System outputs are presented to human users through interpretable interfaces, and human users evaluate recommendations based on professional knowledge, contextual judgment, and value considerations. Feedback is used to update model parameters, knowledge structures, interface strategies, or governance rules.

The update rule is specified as a controlled transition rather than an unspecified semantic gradient:

$$S_{t+1} = \begin{cases} \mathcal{F}_{\theta_t}(S_t, D'_t, F_t), & \text{if the update passes validation,} \\ S_t, & \text{otherwise,} \end{cases} \tag{12}$$

where θ_t identifies the approved model and policy configuration. A rejected model update does not suspend input monitoring: new observations are remapped at the next cycle. Parameter changes require versioning, regression checks, authorization, and a rollback plan. No universal learning-rate interval or stability theorem is assumed. In physical applications, stability must be assessed against the relevant process dynamics and control constraints, rather than inferred from small changes in an internal representation.

Unlike conventional linear AI pipelines, CyTS operates as a recursive cognitive loop. The closed-loop operational process is shown in Figure 3. Conventional systems usually stop after output generation. By contrast, CyTS continuously re-enters the cognitive loop through human evaluation, environmental feedback, and governance adjustment. This recursive mechanism enables long-term co-adaptation between humans and cyber systems. To further illustrate the implementability of CyTS, the iterative human-cyber cognition can be represented as a simplified algorithmic workflow. Algorithm 1 summarizes the major operational steps from entity perception to governance-guided adaptation.

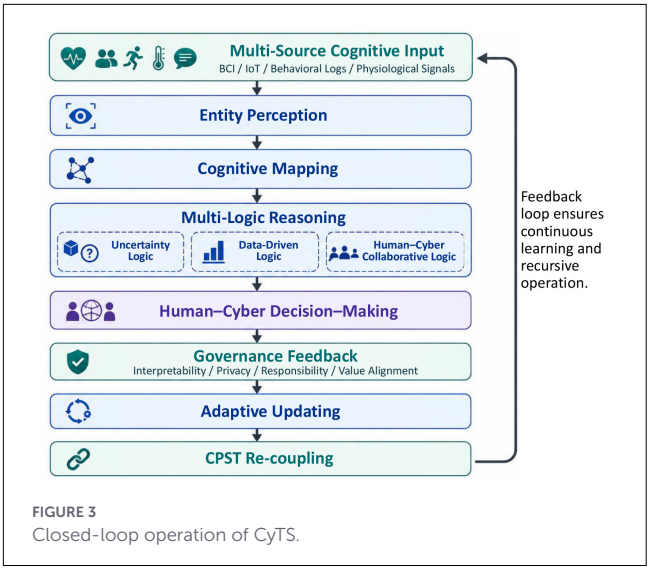


FIGURE 3
Closed-loop operation of CyTS.

The operational process also reflects the coupling structure of CPST. Physical space contributes objective environmental data; social space contributes collective knowledge and institutional rules; cyber space contributes computational modeling capability; and thinking space contributes human cognition, interpretation, and judgment. CyTS coordinates these spaces through cyberlogic-driven organization.

Another key feature is the coexistence of machine-speed processing and human-speed cognition. Cyber models can process multimodal information rapidly and continuously, while humans provide contextual understanding, ethical reasoning, and strategic judgment. CyTS, therefore, does not attempt to eliminate human participation. Instead, it distributes cognitive functions between humans and cyber systems according to their complementary strengths.

The adaptive updating stage is especially important for long-term system evolution. Human feedback can refine knowledge structures, modify reasoning priorities, adjust governance rules, and personalize interaction mechanisms. Over time, CyTS gradually forms domain-specific cognitive adaptation capabilities.

From a systems perspective, CyTS can therefore be understood as a governance-aware cognitive infrastructure rather than an isolated AI model. Its value lies not only in prediction capability but also in continuous cognitive coordination, feedback integration, and organizational stability.

6.2 Cross-domain architectural mapping templates

Although healthcare, education, and transportation differ significantly in operational objectives, environmental constraints, and data modalities, representative CyTS-oriented architectural mappings across these domains share a common Cyberlogic-driven organizational structure. As shown in Table 5 (Ning et al., 2016), CyTS organizes heterogeneous human–cyber systems through a

unified closed-loop architecture consisting of entity, mapping, reasoning, governance, and feedback layers. Table 6 summarizes the functional responsibilities and data exchange relationships among the core components.

Require: Inputs E_t, knowledge K, uncertainty U_t, policies Π_t, deadline, and approved configuration θ_t.
Ensure: An authorized decision or explicit abstention/escalation, with an audit record.
$S_t \leftarrow M(E_t)$; check provenance, timestamps, and units.
Classify the query and select permitted reasoning modules.
$(D_t, \mathcal{A}_t) \leftarrow R(S_t, K, U_t; \Pi_t)$.
Verify assumptions, inference steps, constraints, and evidence status.
if verification fails, required evidence is missing, or the deadline expires **then**
  Abstain or invoke a domain-validated fallback; log the reason.
else
  Request human review when required by risk and authority policy.
  $D'_t \leftarrow G(D_t, HC_t \mid E_x, R_c, H_j, \Pi_t)$.
  Release only an admissible, authorized decision; log evidence and owner.
end if
$F_t \leftarrow$ CollectFeedback(); assess the proposed update.
Apply Equation 12; preserve the configuration and rollback record.
Acquire fresh inputs and advance t; repeat the cycle while the task remains active.

Algorithm 1. Evidence-aware, governance-controlled CyTS cycle.

6.3 Applications of CyTS in different domains

While Table 5 illustrates the shared organizational structure of CyTS across domains, different application scenarios impose distinct governance requirements, human-oversight mechanisms, and operational constraints. Accordingly, CyTS is presented not as a domain-specific AI system with empirically validated performance, but as a conceptual generalized governance-aware human–cyber cognitive architecture adaptable to heterogeneous CPST environments.

Intelligent healthcare represents the most safety-critical CyTS scenario because medical environments require high levels of interpretability, accountability, and human supervision

TABLE 6 Evidence and control contracts at CyTS interfaces.

Interface	Required payload	Failure response
Entity–mapping	Source, timestamp, units, consent/access status, missingness	Reject unauthorized input; flag missing or stale evidence.
Mapping–reasoning	Versioned representation, transformation provenance, uncertainty, relation semantics	Resolve incompatible scales or meanings before inference.
Reasoning–governance	Candidates and $\mathcal{A}_t$: claim type, model, assumptions, provenance, verification status	Withhold unsupported causal claims; expose conflicts and identification failure.
Governance–action	Authorized action or abstention, responsible actor, constraints, deadline, fallback	Escalate or defer if approval, timing, or admissibility checks fail.
Feedback–update	Outcome, action/exposure history, model/policy version, validation and rollback record	Quarantine unvalidated updates; preserve prior approved configuration.

TABLE 5 Unified CyTS structure across application domains.

Dimension	Intelligent healthcare	Smart education	Intelligent transportation
Entity input	Physiological signals, medical images, EEG, wearable data	Learning behavior, eye movement, keyboard logs, classroom interaction	Road sensors, GPS, vehicle data, video streams, weather data
Mapping	Multimodal medical feature extraction	Learning state representation and knowledge tracing	Traffic state representation and risk feature extraction
Reasoning	Diagnostic models, disease risk prediction, medical knowledge graphs	Learning analytics, recommendation models, student state prediction	Traffic prediction, route planning, congestion analysis
Governance	Physician review, clinical responsibility, privacy protection	Teacher intervention, educational fairness, learner privacy	Dispatcher control, safety regulation, responsibility allocation
Feedback	Clinician correction and outcome validation	Teacher feedback and learner performance	Driver response and traffic control outcomes
Core value	Diagnostic reliability and workload reduction	Personalized learning and timely intervention	Safety, efficiency, and congestion reduction

(Chen et al., 2022). Existing AI-assisted diagnostic systems, multimodal clinical decision-support platforms, and medical imaging systems increasingly rely on collaborative human–AI cognition. Within the CyTS perspective, these systems can be conceptually interpreted as governance-aware cognitive augmentation structures in which physicians remain responsible for reviewing recommendations, validating explanations, overriding automated outputs, and ensuring clinical accountability.

As shown in Figure 4, CyTS organizes medical human–AI collaboration as a recursive governance-aware cognitive loop integrating multimodal perception, structured reasoning, physician supervision (Wolpaw and Wolpaw, 2012; Lebedev and Nicolelis, 2006), and adaptive feedback (Tetsuka et al., 2024; Oganesian and Shanechi, 2024). Rather than replacing physician cognition, the architecture supports evidence-based reasoning and reduces information-processing burden while preserving human-centered medical authority.

In smart education, the primary challenge lies in balancing adaptive personalization with fairness, pedagogical responsibility, and learner autonomy. Although recommendation algorithms and learning analytics may offer potential benefits for individualized instruction, educational governance remains fundamentally teacher-centered. Under the CyTS framework, teachers maintain supervisory authority over algorithmic recommendations and contextual educational decisions, thereby preventing purely automated optimization from dominating pedagogical processes.

In intelligent transportation, governance challenges are closely associated with safety regulation, real-time coordination, and responsibility allocation during high-risk operational situations. Existing intelligent traffic-management infrastructures combine predictive AI systems with human dispatching mechanisms to support traffic control and route optimization. Human operators remain responsible for supervisory intervention and safety-critical decision coordination, ensuring that adaptive traffic-management strategies can be designed to remain accountable and controllable under dynamic real-world conditions.

The causal distinction changes the workflow in each domain. In healthcare, a risk prediction is not evidence that a proposed treatment improves outcomes: the record must state the treatment, endpoint, population, confounding assumptions, and available clinical evidence, with clinician review before action. In education, a correlation between a learning resource and achievement does not establish an intervention benefit; evaluation must account for prior attainment and resource assignment. In transportation, an accurate congestion forecast does not establish that signal retiming is safe: candidate actions require a validated dynamics model or controlled simulation and hard traffic constraints. These examples specify requirements rather than report application outcomes.

For a concrete transportation test, let queue length evolve as $x_{t+1} = \max\{0, x_t + a_t - s(x_t, u_t)\}$, where arrivals a_t, signal action u_t, and discharge function s are explicitly defined. The system records which parts of s are measured, assumed, or calibrated; simulates competing actions under arrival uncertainty; checks pedestrian and conflict constraints; and logs the selected action and outcome. A queue equilibrium requires matched arrival and discharge rates under suitable conditions, not merely a repeated observational pattern. This classical testbed can compare alternative relational or equilibrium modules without assuming quantum causation.

Despite domain-specific differences, these scenarios share the same underlying CyTS organizational characteristics: multimodal perception, structured cognitive mapping, multi-logic reasoning coordination, governance-aware human oversight, and recursive feedback adaptation. This structural consistency suggests that Cyberlogic can serve as an organizational foundation for human–cyber cognitive augmentation across heterogeneous CPST scenarios.

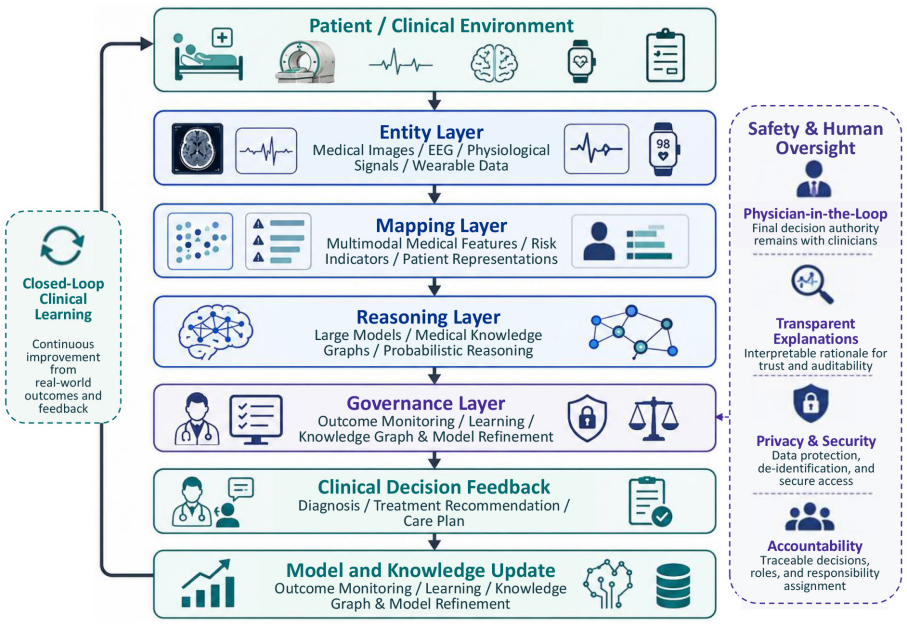


FIGURE 4
Cyberlogic-driven CyTS operation in intelligent healthcare.

7 Evaluation framework and governance challenges

While the foregoing applications are conceptual illustrations, practical deployment of CyTS in real-world settings would require systematic evaluation. Since CyTS directly influences human cognition and decision-making, its evaluation should not be limited to model accuracy alone (Burton et al., 2024). Instead, CyTS should be regarded as a governance-aware human–cyber cognitive system integrating reasoning quality, interpretability, collaboration effectiveness, adaptability, and preservation of human agency (Devlin et al., 2019). To improve operational applicability, each evaluation dimension should include measurable indicators and corresponding assessment methods. Table 7 summarizes a governance-aware evaluation framework for CyTS systems.

7.1 Causal and resource evaluation protocol

A first prototype should compare a common predictive/LLM backbone in four conditions: baseline assistance, governance logging only, governance plus explicit causal verification, and an exploratory equilibrium module where an executable implementation exists. Keep data, hardware, model versions, prompts, and action constraints fixed, and use paired task instances with repeated runs. Human-facing comparisons require controlled assignment and a prespecified analysis of workload and automation bias. Simulator interventions with known ground truth and formal causal benchmarks complement observational evaluations; neither replaces prospective domain validation.

For intervention queries, score correct identification or justified non-identification, effect-estimation error when ground truth exists, and sensitivity to hidden-confounder or mechanism-shift scenarios. Independently audit whether reported evidence corresponds to executed reasoning. Include failures and abstentions in the denominator when reporting coverage, so that selective refusal cannot create an artificial gain in reliability. Equilibrium modules must additionally state solution existence, multiplicity, stability conditions, and behavior away from equilibrium.

Measure whole-pipeline energy per completed task and per accepted correct decision, alongside the coverage and safety results. State the measurement boundary, hardware utilization, batch size, output length, idle-power treatment, and repeated-run uncertainty. Report latency and human effort separately. The aim is to determine whether selective verification avoids enough unnecessary computation to offset its own overhead; CyTS currently provides no measured energy advantage.

7.2 Evidence boundaries and research agenda

CyTS currently offers an organizational specification, not a validated cognitive theory or a deployed causal engine. The proposed interfaces improve auditability only if implementations faithfully record computation and independently check evidence. They do not resolve all causal identification problems, guarantee privacy, or prove closed-loop stability. Causal validity, normative acceptability, and human authorization are separate conditions: an effective intervention may still be inadmissible.

Equilibrium-based and EMTI-inspired components remain optional research extensions. Before they influence deployed actions, they should provide formal semantics, reproducible implementations, measurable predictions, and comparisons with classical alternatives under the same resources. A claim involving physical quantum effects additionally requires an operationally specified physical system and suitable measurements. Architectural compatibility or a network analogy alone cannot meet that requirement. The near-term priority is a conventional prototype that tests the evidence contracts and governance gates; broader foundational claims remain open to separate investigation.

However, achieving satisfactory performance across these dimensions remains challenging in real-world deployments. In particular, interpretability, human agency preservation, privacy

TABLE 7 Evaluation dimensions for CyTS systems.

Evaluation dimension	Core question	Representative indicators	Measurement protocol/instrument
Cognitive augmentation	Does the system improve human cognitive performance?	Task accuracy, completion time, cognitive workload reduction	Human-only vs. CyTS-assisted comparison; NASA-TLX
Collaborative gain	Does human–cyber collaboration outperform either side alone?	Error correction rate, collaborative efficiency	Human-only vs. AI-only vs. human–cyber comparison
Interpretability	Can users understand system outputs and reasoning processes?	Explanation usefulness, uncertainty awareness	SHAP/LIME explanation analysis; user trust evaluation
Governance effectiveness	Are risks and ethical constraints effectively controlled?	Auditability, bias monitoring, privacy compliance	Governance audit and fairness evaluation
Adaptability	Can the system improve through feedback?	Feedback utilization rate, update stability	Longitudinal adaptive-performance tracking
Human agency preservation	Does the system preserve human decision authority?	Override frequency, automation-bias reduction	Human intervention and override analysis

protection, and responsibility allocation constitute the most critical governance concerns affecting trustworthy human–cyber cognitive augmentation systems. The major governance challenges of CyTS can be summarized as follows:

1. **Interpretability, trust, and transparency:** since CyTS directly influences human cognition and decision-making, users must understand not only system recommendations but also the supporting evidence, reasoning process, and associated uncertainty. Consequently, explainability should be embedded within the Governance layer rather than treated as a *post-hoc* add-on (Bubeck et al., 2023; Ji et al., 2023; Hogan et al., 2021; Rudin, 2019).
2. **Human agency preservation and automation bias:** as CyTS systems become increasingly capable, users may become overly dependent on automated recommendations, resulting in reduced professional judgment and automation bias (Silver et al., 2016). Maintaining meaningful human control therefore becomes a fundamental governance requirement. CyTS addresses this issue through uncertainty visualization, counterfactual explanations, decision confirmation mechanisms, and human override capabilities that preserve human agency throughout the cognitive loop.
3. **Privacy and cognitive data protection:** CyTS processes highly sensitive information, including cognitive states, physiological signals, behavioral patterns, and decision preferences. Governance-aware privacy mechanisms such as anonymization, differential privacy (Dwork, 2006), access control, and secure data management are therefore essential. Privacy protection should be integrated directly into the system architecture rather than imposed as an external compliance mechanism.
4. **Responsibility allocation and accountable governance:** responsibility allocation becomes increasingly complex because decisions emerge from interactions between human cognition and cyber recommendations (Topol, 2019; Kamar, 2016; Esteva et al., 2017). Inspired by contemporary AI governance principles and the notion of meaningful human control, CyTS incorporates human-in-the-loop supervision, decision traceability, risk-tier governance, and human override mechanisms. Under this structure, developers, deployers, institutions, and end users assume differentiated yet traceable responsibilities, while humans retain final authority in safety-critical situations.

8 Conclusion

This paper proposed CyTS as a governance-centric organizational architecture for human–cyber cognition in CPST environments. Its entity–mapping–reasoning–governance structure connects heterogeneous representations, reasoning evidence, human judgment, and controlled adaptation. The framework distinguishes logical validity and predictive association from intervention-relevant causal support, and specifies evidence records, component interfaces, and decision gates that preserve these distinctions. Healthcare, education, and transportation scenarios illustrate how those requirements affect decisions; they do not demonstrate effectiveness or deployment readiness.

The theoretical scope includes structural causal models and equilibrium-based causal perspectives, while brain–universe analogies and EMTI are situated as exploratory directions with distinct evidentiary requirements. CyTS does not claim that entity modeling, symbolic inference, or a proposed quantum account is individually sufficient for cognition. Future work should implement the specified contracts, test transparency and decision-quality hypotheses under controlled interventions and distribution shifts, and measure energy and latency at matched safety and coverage. Any claimed benefit of a new causal formalism must be established through reproducible comparisons rather than inferred from architectural integration.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

RC: Writing – original draft, Methodology, Conceptualization. LZ: Methodology, Writing – original draft. JD: Writing – review & editing. HN: Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., et al. (2019). "Guidelines for human-AI interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems* (Glasgow), 1–13. doi: 10.1145/3290605.3300233

Baltrušaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal machine learning: a survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 423–443. doi: 10.1109/TPAMI.2018.2798607

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). "On the dangers of stochastic parrots: can language models be too big?" in *Proceedings of the ACM conference on fairness, accountability, and transparency* (Canada: Virtual Event), 610–623. doi: 10.1145/3442188.3445922

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., et al. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv* [Preprint]. arXiv:2004.07213. doi: 10.48550/arXiv.2004.07213

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., et al. (2023). Sparks of artificial general intelligence: early experiments with GPT-4. *arXiv* [Preprint]. arXiv:2303.12712. doi: 10.48550/arXiv.2303.12712

Burton, J. W., Lopez-Lopez, E., Hechtlinger, S., Rahwan, Z., Aeschbach, S., Bakker, M. A., et al. (2024). How large language models can reshape collective intelligence. *Nat. Hum. Behav.* 8, 2183–2195. doi: 10.1038/s41562-024-01959-9

Chang, T. A., and Bergen, B. K. (2024). Language model behavior: a comprehensive survey. *Comput. Linguist.* 50, 293–349. doi: 10.1162/coli_a_00492

Chen, H., Gomez, C., Huang, C. M., and Unberath, M. (2022). Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. *NPJ Digit. Med.* 5:156. doi: 10.1038/s41746-022-00699-2

Collins, K. M., Sucholutsky, I., Bhatt, U., Chandra, K., Wong, L., Lee, M., et al. (2024). Building machines that learn and think with people. *Nat. Hum. Behav.* 8, 1851–1863. doi: 10.1038/s41562-024-01991-9

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. (2019). Hybrid intelligence. *Bus. Inf. Syst. Eng.* 61, 637–643. doi: 10.1007/s12599-019-00595-2

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT* (Minneapolis), 4171–4186. doi: 10.18653/v1/N19-1423

Dhelim, S., Ning, H., Cui, S., Ma, J., Huang, R., and Wang, K. I.-K. (2020). Cyberentity and its consistency in the cyber-physical-thinking hyperspace. *Comput. Electr. Eng.* 81:106506. doi: 10.1016/j.compeleceng.2019.106506

Durstewitz, D., Koppe, G., and Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Mol. Psychiatry* 24, 1583–1598. doi: 10.1038/s41380-019-0365-9

Dwork, C. (2006). "Differential privacy," in *Proceedings of the 33rd international colloquium on automata, languages and programming* (Venice), 1–12. doi: 10.1007/11787006_1

Edelman, B. J., Meng, J., Suma, D., Zurn, C., Nagarajan, E., Baxter, B. S., et al. (2019). Noninvasive neuroimaging enhances continuous neural tracking for robotic control. *Sci. Robot.* 4:eaaw6844. doi: 10.1126/scirobotics.aaw6844

Edelman, B. J., Zhang, S., Schalk, G., Brunner, P., Muller-Putz, G., Guan, C., et al. (2025). Non-invasive brain-computer interfaces: state of the art and trends. *IEEE Rev. Biomed. Eng.* 18, 26–49. doi: 10.1109/RBME.2024.3449790

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056

Flesher, S. N., Downey, J. E., Weiss, J. M., Hughes, C. L., Herrera, A. J., Tyler-Kabara, E. C., et al. (2021). A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science* 372, 831–836. doi: 10.1126/science.abd0380

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People–an ethical framework for a good AI society. *Minds Mach.* 28, 689–707. doi: 10.1007/s11023-018-9482-5

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Gao, X., Wang, Y., Chen, X., and Gao, S. (2021). Interface, interaction, and intelligence in generalized brain-computer interfaces. *Trends Cogn. Sci.* 25, 671–684. doi: 10.1016/j.tics.2021.04.003

Geirhos, R., Jacobsen, J. H., Michaelis, C., et al. (2020). Shortcut learning in deep neural networks. *Nat. Mach. Intell.* 2, 665–673. doi: 10.1038/s42256-020-00257-z

Gilbert, S., Kather, J. N., and Hogan, A. (2024). Augmented non-hallucinating large language models as medical information curators. *NPJ Digit. Med.* 7:143. doi: 10.1038/s41746-024-01081-0

Glickman, M., and Sharot, T. (2025). How human-AI feedback loops alter human judgment. *Nat. Hum. Behav.* 9, 1–10. doi: 10.1038/s41562-024-02077-2

Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutiérrez, C., et al. (2021). Knowledge graphs. *ACM Comput. Surv.* 54:71. doi: 10.1145/3447772

Huang, S. C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., and Chaudhari, A. S. (2023). Self-supervised learning for medical image classification. *NPJ Digit. Med.* 6, 1–16. doi: 10.1038/s41746-023-00811-0

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., et al. (2023). Survey of hallucination in natural language generation. *ACM Comput. Surv.* 55:248. doi: 10.1145/3571730

Jin, Z., Chen, Y., Leeb, F., Gresele, L., Kamal, O., Lyu, Z., et al. (2023). CLadder: assessing causal reasoning in language models. *arXiv* [Preprint]. arXiv:2312.04350. doi: 10.52202/075280-1353

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392

Kamar, E. (2016). "Directions in hybrid intelligence: complementing AI systems with human intelligence," in *Proceedings of the 25th international joint conference on artificial intelligence* (New York, NY), 4070–4073.

Kosinski, M. (2017). Facial recognition technology can expose political orientation. *Nat. Hum. Behav.* 1, 1–3. doi: 10.1038/s41598-020-79310-1

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behav. Brain Sci.* 40:e253. doi: 10.1017/S0140525X16001837

Lebedev, M. A., and Nicolelis, M. A. (2006). Brain-machine interfaces: past, present and future. *Trends Neurosci.* 29, 536–546. doi: 10.1016/j.tins.2006.07.004

Lorach, H., Galvez, A., Spagnolo, V., Martel, F., Karakas, S., Intering, N., et al. (2023). Walking naturally after spinal cord injury using a brain-spine interface. *Nature* 618, 126–133. doi: 10.1038/s41586-023-06094-5

Lotte, F., Larrue, F., and Mühl, C. (2013). Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. *Front. Hum. Neurosci.* 7:568. doi: 10.3389/fnhum.2013.00568

Luccioni, A. S., Jernite, Y., and Strubell, E. (2024). "Power hungry processing: watts driving the cost of AI deployment?," in *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*, 85–99. doi: 10.1145/3630106.3658542

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data Soc.* 3, 1–21. doi: 10.1177/2053951716679679

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236

Mooij, J. M., Janzing, D., and Schölkopf, B. (2013). From ordinary differential equations to structural causal models: the deterministic case. *arXiv* [Preprint] arXiv:1304.7920.

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., et al. (2023). Foundation models for generalist medical artificial intelligence. *Nature* 616, 259–265. doi: 10.1038/s41586-023-05881-4

Morey, D. A., Rayo, M. F., and Woods, D. D. (2025). Empirically derived evaluation requirements for responsible AI augmentation in safety-critical settings. *NPJ Digit. Med.* 8:223. doi: 10.1038/s41746-025-01784-y

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y. (2011). "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning* (Bellevue, WA), 689–696.

Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L., and Konstan, J. A. (2014). "Exploring the filter bubble: the effect of using recommender systems," *in Proceedings of WWW* (Seoul), 677–686. doi: 10.1145/2566486.2568012

Ning, H. (2016). Cybermatics: cyber-physical-social-thinking hyperspace for the internet of things. *Future Gener. Comput. Syst.* 56, 504–512. doi: 10.1016/j.future.2015.07.012

Ning, H., Li, Q., Wei, D., Liu, H., and Zhu, T. (2017). Cyberlogic paves the way from cyber philosophy to cyber science. *IEEE Internet Things J.* 4, 783–790. doi: 10.1109/JIOT.2017.2666798

Ning, H., Lin, Y., Wang, W., Wang, H., Shi, F., Zhang, X., et al. (2023). Cyberology: cyber-physical-social-thinking spaces-based discipline and interdiscipline hierarchy for metaverse (general cyberspace). *IEEE Internet Things J.* 10, 4420–4430. doi: 10.1109/JIOT.2022.3217821

Ning, H., and Liu, H. (2015). Cyber-physical-social-thinking space based science and technology framework for the internet of things. *Sci. China Inf. Sci.* 58, 1–19. doi: 10.1007/s11432-014-5209-2

Ning, H., Liu, H., and Yang, L. (2016). Cyber-physical-social systems: the state of the art and perspectives. *IEEE Trans. Syst. Man Cybern. Syst.* 46, 465–478. doi: 10.1109/TCSS.2018.2861224

Ning, H., Ye, X., Bouras, M. A., and Wei, D., and Daneshmand, M. (2018). General cyberspace: cyberspace and cyber-enabled spaces. *IEEE Internet Things J.* 5, 1843–1856. doi: 10.1109/JIOT.2018.2815535

Oganesian, L. L., and Shanechi, M. M. (2024). Brain-computer interfaces for neuropsychiatric disorders. *Nat. Rev. Bioeng.* 2, 653–670. doi: 10.1038/s44222-024-00177-2

Pahud de Mortanges, A., Luo, H., Shu, S. Z., Kamath, A., Suter, Y., Shelan, M., et al. (2024). Orchestrating explainable artificial intelligence for multimodal and longitudinal data in medical imaging. *NPJ Digit. Med.* 7:195. doi: 10.1038/s41746-024-01190-w

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. New York, NY: Penguin Press.

Patrick-Krueger, K. M., Burkhart, I., and Contreras-Vidal, J. L. (2025). The state of clinical trials of implantable brain-computer interfaces. *Nat. Rev. Bioeng.* 3, 1–15. doi: 10.1038/s44222-024-00239-5

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *Commun. ACM* 61, 64–72. doi: 10.1145/3159652.3176182

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., et al. (2019). Machine behaviour. *Nature* 568, 477–486. doi: 10.1038/s41586-019-1138-y

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., GI Genius CADx Study Group, et al. (2022). Experimental evidence of effective human-AI collaboration in medical decision-making. *Sci. Rep.* 12:14952. doi: 10.1038/s41598-022-18751-2

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 16:051001. doi: 10.1088/1741-2552/ab260c

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. doi: 10.1038/s42256-019-0048-x

Schierman, J. D., DeVore, M. D., Richards, N. D., and Clark, M. A. (2020). Runtime assurance for autonomous aerospace systems. *J. Guid. Control Dyn.* 43, 2205–2217. doi: 10.2514/1.G004862

Shi, Q., Liu, F., Li, H., Li, G., Shi, L., and Zhao, R. (2025). Hybrid neural networks for continual learning inspired by brain mechanisms. *Nat. Commun.* 16, 1–13. doi: 10.1038/s41467-025-56405-9

Shneiderman, B. (2020). Human-centered artificial intelligence: reliable, safe & trustworthy. *Int. J. Hum. Comput. Interact.* 36, 495–504. doi: 10.1080/10447318.2020.1741118

Shrestha, R., Leinonen, J., Zavgorodniaia, A., Hellas, A., and Edwards, J. (2022). "Pausing while programming: insights from keystroke analysis," in *Proceedings of the 2022 ACM conference on international computing education research* (Pittsburgh, PA), 41–51. doi: 10.1145/3510456.3514146

Silver, D., Huang, A., Maddison, CJ., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Sundrani, S., Chen, J., Jin, B. T., Abad, Z. S. H., Rajpurkar, P., and Kim, D. (2023). Predicting patient decompensation using multimodal data. *NPJ Digit. Med.* 6, 1–12. doi: 10.1038/s41746-023-00803-0

Tetsuka, H., Gobbi, S., Hatanaka, T., Pirrami, L., and Shin, S. R. (2024). Wirelessly steerable bioelectronic neuromuscular robots adapting neurocardiac junctions. *Sci. Robot.* 9:eado0051. doi: 10.1126/scirobotics.ado0051

Tonekaboni, S., Joshi, S., McCradden, M. D., and Goldenberg, A. (2019). "What clinicians want: contextualizing explainable machine learning for clinical end use," in *Proceedings of the machine learning for healthcare conference* (Ann Arbor MI), 1–12.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 44–56. doi: 10.1038/s41591-018-0300-7

Turpin, M., Michael, J., Perez, E., and Bowman, S. R. (2023). Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting. *Adv. Neural Inf. Process. Syst.* 36, 74952–74965. doi: 10.52202/075280-3275

Vaccaro, M., Almaatouq, A., and Malone, T. (2024). When combinations of humans and AI are useful. *Nat. Hum. Behav.* 8, 1–10. doi: 10.31234/osf.io/mk46p

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in neural information processing systems* (Long Beach, CA), 5998–6008.

Vazza, F., and Feletti, A. (2020). The quantitative comparison between the neuronal network and the cosmic web. *Front. Phys.* 8:525731. doi: 10.3389/fphy.2020.525731

Wang, G., Tang, Y., Luo, X., Lu, S., Zhou, Y., Lu, Y., et al. (2025). Printed sensing human-machine interface with adaptive response. *Sci. Adv.* 11:eadw3725. doi: 10.1126/sciadv.adw3725

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., et al. (2023). A high-performance speech neuroprosthesis. *Nature* 620, 1031–1036. doi: 10.1038/s41586-023-06377-x

Wilson, G. H., Stein, E. A., Kamdar, F., Avansino, D. T., Pun, T. K., Gross, R., et al. (2025). Long-term recalibration of intracortical brain-computer interfaces. *Nat. Biomed. Eng.* 9, 1–12. doi: 10.1038/s41551-025-01536-z

Wolpaw, J. R., and Wolpaw, E. W. (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford: Oxford University Press.

Yang, J., Eyre, D. W., Lu, L., and Clifton, D. A. (2023). Interpretable machine learning for antimicrobial resistance prediction. *Commun. Med.* 3, 1–12. doi: 10.1038/s44259-023-00015-2

Zeng, J., Yang, L. T., Lin, M., Ning, H., and Ma, J. (2020). A survey: cyber-physical-social systems and their system-level design methodology. *Future Gener. Comput. Syst.* 105, 489–507. doi: 10.1016/j.future.2016.06.034

Zhang, W.-R. (2024). Global realism with bipolar strings: from bell test to real-world causal-logical quantum gravity and brain-universe similarity for entangled machine thinking and imagination. *Information* 15:456. doi: 10.3390/info15080456

Zhang, W.-R., and Peace, K. E. (2014). Causality is logically definable–toward an equilibrium-based computing paradigm of quantum agent and quantum intelligence (QAQI) (survey and research). *J. Quantum Inf. Sci.* 4, 227–268. doi: 10.4236/jqis.2014.44021