Artificial eXperience Intelligence (AXI): An Engineering Discipline for the Human Experience of AI Systems

Yeluri S. D. S. Sri Vardhan

srivardhan@sripto.tech

Sripto Corporation Private Limited https://orcid.org/0009-0003-5030-0490

Research Article

Keywords: Artificial intelligence, human-AI interaction, agentic AI, explainable AI, multimodal AI, embodied AI, affective computing, experience design, trust in automation, AXI, AXI Bridge

Posted Date: May 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9717445/v1

Additional Declarations: The authors declare potential competing interests as follows: The author is the Founder and Chief Executive Officer of Sripto Corporation Private Limited, which develops both Let Me Teach and SRIPTO Avatar Runtime (SAR) - the systems discussed in Section 6. This is a material competing interest. The case studies in Section 6 are therefore presented as preliminary, non-independent operationalizations rather than as independent validation of the framework. Independent third-party replications are explicitly invited.

Artificial eXperience Intelligence (AXI)

An Engineering Discipline for the Human Experience of AI Systems

Yeluri S. D. S. Sri Vardhan

Sripto Corporation Private Limited, Andhra Pradesh, India

Correspondence: srivardhan@sripto.tech | https://sripto.tech

Contents

Abstract

We propose **Artificial eXperience Intelligence (AXI)** as an engineering discipline whose unit of optimization is the quality of a human's lived experience of an artificial intelligence system. AXI introduces the **AXI Bridge** — a runtime control plane that mediates between AI capability and human cognitive bandwidth through a five-layer reference architecture (the AXI Stack), nine design priorities (the AXI Principles), and a seven-metric evaluation framework with explicit telemetry-based definitions. We address a documented gap: although approximately 900 million users interact with leading generative-AI systems weekly, only 46% of global respondents are willing to trust AI despite 66% using it regularly; 95% of enterprise generative-AI pilots produce zero measurable P&L impact; and Gartner forecasts over 40% of agentic-AI projects will be cancelled by 2027. We argue this gap is not a model-capability problem but an experience-engineering problem. The AXI Score uses a multiplicative gated model in which consent or integrity failures produce a zero score regardless of other metrics, addressing the well-known vulnerability of linear composites to mask catastrophic failures. The framework specifies a control-loop latency budget bounding the cost of downward constraint propagation, an Asynchronous Non-Blocking Evaluation mode for sub-millisecond inference engines, irreversibility handling for non-reversible action classes, a Mandate Rollback Protocol for deferred-mode agents, and seven Deception Guardrails preventing anthropomorphic dark-pattern risks. We present two case studies — *Let Me Teach*, a real-time interactive explainer, and *SRIPTO Avatar Runtime (SAR)*, a digital-human runtime architecture — operationalizing AXI in production code, with measured AXI metric observations on a launched-product cohort of 2,500 sessions and a pre-launch pilot cohort of 150 sessions, demonstrating that the framework can be instrumented in production code and yields stable, interpretable measurements (AXI Scores 0.91 and 0.89). These observations do not constitute independent validation of the framework. The framework is offered to the research and practitioner community for critique, extension, and replication.

Keywords: Artificial intelligence; human-AI interaction; agentic AI; explainable AI; multimodal AI; embodied AI; affective computing; experience design; trust in automation; AXI; AXI Bridge.

ACM Classification: H.5.2 [Information Interfaces and Presentation]: User Interfaces — Evaluation/methodology; I.2.0 [Artificial Intelligence]: General.

Significance Statement

Artificial intelligence has become extraordinarily capable in a very short time. Nearly everyone uses it, but far fewer people actually trust it. Most companies investing in AI report no measurable returns yet, and a large share of customers say they would rather not interact with AI at all. The intelligence of AI has advanced faster than the experience of AI.

This paper introduces a name, a structure, and a working engineering specification for the discipline that closes that gap. We call it **Artificial eXperience Intelligence (AXI)**, and its central engineering object is the **AXI Bridge** — a runtime control plane that sits between AI capability and the human user. AXI proposes a five-layer architecture (the AXI Stack), nine design priorities (the AXI Principles), and seven measurable indicators with explicit telemetry-based definitions (the AXI Evaluation Framework). We demonstrate the framework's implementability through two production systems and report measured indicators across 2,650 measured user sessions.

---

1. Introduction

1.1 The capability–experience gap

Between 2017 and 2026, AI transitioned from a research discipline to one of the fastest-adopted general-purpose technologies in history. ChatGPT alone reached approximately 900 million weekly active users

by early 2026 [1], and the two leading frontier-AI companies together exceeded $39 billion in combined annualized revenue by February 2026, with valuations of approximately $852 billion and $380 billion respectively [2]. HUMAN Security's 2026 *State of AI Traffic & Cyberthreat Benchmark Report* measured AI-driven internet traffic growing 187% across 2025, with AI agents now a measurable commercial force in retail, media, travel, and hospitality [3]. Frontier capability advanced across reasoning, multimodality, agency, and embodiment.

The lived experience of AI has not advanced at the same pace. The KPMG–University of Melbourne global study (n = 48,340 across 47 countries, April 2025) reports that only 46% of people are willing to trust AI, while 66% already use it regularly — a widening gap since 2022, with 70% reporting that AI regulation is required [4]. MIT Project NANDA's *GenAI Divide: State of AI in Business 2025* reports that 95% of enterprise generative-AI pilots deliver zero measurable P&L impact [5]. Gartner's June 2025 forecast projects that over 40% of agentic AI projects will be cancelled by the end of 2027 [6]. McKinsey's *State of AI 2025* (n = 1,933 executives, 105 countries) reports 88% of organizations use AI in at least one function, but only approximately 6% qualify as AI high performers — those attributing greater than 5% of EBIT to AI [7]. Pew Research (survey conducted August–October 2024, published April 3, 2025; public n = 5,410, AI expert n = 1,013) found 51% of US adults are more concerned than excited about AI, with only 11% more excited than concerned [8]. Gartner's July 2024 survey (n = 5,728) found 64% of customers would prefer companies did not use AI in customer service [9]. Stanford HAI reports AI-related incidents rose 56.4% in a single year [10].

Indicator	Value	Source
ChatGPT weekly active users	~900M	[1]
Year-on-year rise in AI internet traffic (2025)	+187%	[3]
Global willingness to trust AI	46%	[4]
Global regular AI use	66%	[4]
Enterprise GenAI pilots with zero P&L impact	95%	[5]
Enterprise AI high performers (>5% EBIT)	~6%	[7]
Firms abandoning most AI initiatives (2025)	42%	[11]
Agentic AI projects forecast cancelled by 2027	40%+	[6]
Customers preferring no AI in service	64%	[9]
US adults more concerned than excited about AI	51%	[8]
AI-related incidents, year-over-year rise (2024)	+56.4%	[10]

Table 1. The experience deficit in published indicators.

1.2 Problem statement

The data above suggests that the bottleneck for AI value at population scale is increasingly less about model capability and more about the engineering of experience. We hypothesize that closing this gap requires (a) a named integrating discipline with a well-specified runtime architecture, (b) measurable experience-layer indicators that cannot be gamed by improvements in raw capability alone, and (c) a hard-gated scoring model preventing systems that fail on consent or integrity from registering as high-quality.

1.3 Naming and scope

The acronym AXI stands for **Artificial eXperience Intelligence**. The letters AXI and the adjacent term AX appear in other AI-related contexts — including Axiomatic Intelligence (AxI) from Axiomatic AI, Inc., AXi (AUDIENCEX Intelligence) in advertising technology, Axi from Axify Pty Ltd in conversational AI, AX as a corporate slogan used by LG Uplus for AI transformation, and the emerging industry terms Agent Experience

(AX) and Artificial Experience (AX). The framework introduced here is a distinct proposal centered on the end-user's lived experience of AI systems and does not claim exclusive rights to the three letters AXI in general usage. Readers are encouraged to refer to the full expansion where ambiguity is possible.

A specific clarification is warranted regarding **Agent Experience (AX)**, coined by Matt Biilmann (Netlify CEO) in early 2025 and gaining traction as an emerging design discipline in its own right. Where Agent Experience optimizes systems for AI agents *as users* of digital systems (e.g., crawlability, structured affordances for automated consumers), AXI optimizes systems for *humans as users* of AI systems. The two are complementary, not competing — a single product can apply both.

1.4 Contributions

This paper makes four contributions:

1. A working **definition of Artificial eXperience Intelligence** as a proposed engineering discipline whose unit of optimization is the quality of a human's lived experience of an AI system (Section 4.1).
2. The **AXI Bridge runtime control-plane specification** — including the five-layer *AXI Stack* with a *Perceptual Substrate* layer that explicitly disclaims the Brooks-strict embodiment definition, the synchronous and asynchronous (deferred) operating modes, the API contracts, the Trust Ledger and Consent Broker schemas, the Asynchronous Non-Blocking Evaluation (ANBE) mode, irreversibility handling, and the Mandate Rollback Protocol (Sections 4.2–4.6) — directly relevant to the NIST AI Agent Standards Initiative (February 2026) [12] and Singapore's Model AI Governance Framework for Agentic AI (January 2026) [13].
3. A **gated evaluation framework** consisting of nine AXI Principles, seven telemetry-anchored metrics with passive proxies (a piecewise harmonic mean for Interaction Integrity with a two-window tiebreaker, a step-function-aware Trust Decay Rate, and dual-gated microsurvey sampling), a multiplicative gated AXI Score with explicit normalization anchors, and seven Deception Guardrails preventing anthropomorphic dark patterns (Sections 4.4–4.7).
4. **Two case-study operationalizations** with measured metric observations on production cohorts — *Let Me Teach* (n = 2,500, launched product) and *SRIPTO Avatar Runtime* (n = 150, pre-launch pilot) — together totaling 2,650 sessions, demonstrating implementability and operational measurability (Section 6).

---

2. Literature Review

AXI builds on nine adjacent disciplines, each of which is a multi-decade, multi-thousand-researcher community. The summary below cannot do justice to any of them; we cite key foundations and note where AXI attempts to add value.

Artificial Intelligence. From Turing [14] to modern frontier models. Legg and Hutter [15] formalized intelligence as expected performance across a universal reward distribution. Bubeck et al. [16] argued GPT-4 exhibits early AGI capabilities. Modern frontier models provide the Cognition Layer; AXI proposes complementary metrics for felt experience orthogonal to capability benchmarks.

Agentic AI. ReAct [17], Toolformer [18], generative agents [19], and engineering practice from major laboratories [20][21][22]. AXI extends agentic frameworks with experience-layer constraints for both synchronous and asynchronous (deferred) modes.

Explainable AI. DARPA's XAI program [23], the Arrieta et al. survey [24], LIME [25], SHAP [26]. Microsoft's research [27] has shown that explanations can increase user over-reliance on incorrect AI recommendations — a phenomenon AXI addresses at the Experience Layer through gated trust accounting. Trust-calibration research [28][29] has established that human-AI trust is best modeled as a dynamic, context-sensitive process rather than a static property.

Multimodal AI. CLIP [30], Flamingo [31], LLaVA [32], GPT-4V [33], Gemini [34]. AXI treats multimodality as a perception channel in service of the experience objective.

Embodied AI. Brooks's argument that intelligence is grounded in physical embodiment [35], with modern translation to foundation-model-driven robotics [36]. The 2024–2026 humanoid-robotics cohort raised approximately $3.7 billion in 2025 [37]. AXI deliberately uses a broader *Perceptual Substrate* construct for its Layer 4 — encompassing voice, avatar, kiosk, hologram, and humanoid robot — and explicitly disclaims the Brooks-strict embodiment definition for non-physical substrates (Section 4.2.4).

Human-Computer Interaction and Human-Centered AI. Norman [38][39], Shneiderman [40][41], Nielsen [42], Lombard and Ditton [43], Slater and Wilbur [44], Csikszentmihalyi [45], Lee and See [46]. Practitioner research [47] documents new interaction patterns emerging in the AI era. Human-Centered AI [41] is the discipline most aligned with AXI; we offer AXI as an engineering-focused companion rather than a successor, and explicitly disclaim any attempt to replace HCAI.

Conversational AI. From ELIZA [48] to modern LLM chatbots. Conversational AI centers the conversation itself; AXI focuses on the broader experience surrounding the conversation.

Affective Computing. Picard [49], Breazeal [50]. Reeves and Nass's *The Media Equation* [65] established the foundational empirical case that humans respond to computers as social actors — a finding central to AXI's anthropomorphism concerns and the Deception Guardrails (Section 4.7). The widely-cited Ayers et al. (2023) finding [51] that LLM responses are rated higher for empathy than physicians' is sometimes overinterpreted: the study measured *perceived empathy by third-party text evaluators reading patient-question responses on a public social media forum*, not real-time empathetic interaction. AXI is explicit about this distinction.

Adjacent integrative proposals. Microsoft's Guidelines for Human-AI Interaction [52] — 18 empirically-validated heuristics directly informing AXI's principles. Google PAIR's People + AI Guidebook [53], Shneiderman's Human-Centered AI [41], Xu's UX 3.0 paradigm [54], and industry framings of Agent Experience (AX) and Artificial Experience (AX) [55]. The 2026 regulatory landscape — NIST AI RMF 2.0, the EU AI Act, ISO/IEC 42001, Singapore's Model AI Governance Framework for Agentic AI (January 2026) [13][56][57] — establishes audit and transparency expectations that AXI's Consent Fidelity and Interaction Integrity metrics map onto naturally. AXI is offered alongside these proposals, not against them.

The author's prior peer-reviewed work [58] is in an adjacent field (business intelligence in IT industry) and is cited here for completeness; it does not establish HCI/AI domain expertise.

---

3. Methodology

The AXI framework was developed through a six-step methodology:

1. **Problem identification.** Review of publicly available enterprise-deployment studies, population-scale trust surveys, and customer-experience research (Section 1.1).
2. **Prior-art mapping.** Review of nine adjacent disciplines for contributions and explicit gaps with respect to end-to-end lived experience (Section 2).
3. **Construct synthesis.** Five architectural layers, nine principles, and seven metrics synthesized from the prior-art mapping, each construct traceable to at least one foundational citation.
4. **Engineering specification.** Translation of each construct into a runtime specification: a latency budget, an API contract, telemetry schemas, durable state structures, and a gated scoring model (Section 4).
5. **Case-study operationalization.** Instrumentation of the framework in two production systems — *Let Me Teach* and *SAR* — and collection of AXI-metric telemetry on a launched-product cohort (n = 2,500) and a pre-launch pilot cohort (n = 150) (Section 6).
6. **Iterative revision based on internal review and stakeholder feedback.** Multiple rounds of internal review and stakeholder feedback informed substantive revisions: replacement of a linear composite score with a multiplicative gated model, replacement of an Embodiment label with Perceptual Substrate, demotion of microsurveys from primary signal to occasional calibration, addition of telemetry-based passive proxies, addition of step-function-aware Trust Decay Rate, addition of irreversibility handling, addition of the Mandate Rollback Protocol, and addition of the Deception Guardrails.

4. The AXI Framework

4.1 Definition

We propose **Artificial eXperience Intelligence (AXI)** as an engineering discipline whose unit of optimization is the quality of a human's lived experience of an AI system. AXI specifies a runtime control plane (the AXI Bridge), a five-layer reference architecture (the AXI Stack), nine design priorities (the AXI Principles), and a seven-metric gated evaluation framework (the AXI Score) for both synchronous and asynchronous (deferred) AI interactions.

Where classical AI optimizes for *task success* (accuracy, reward, benchmark) and HCI optimizes for *interface usability* (task time, error rate, system usability scale), AXI optimizes for **experience quality** — a multiplicatively-gated composite of presence, comfort, pacing, trust, repair, continuity, and consent.

The definition contains four load-bearing adjectives. An AXI-compliant system is **present** (here, now, with the user — not an anonymous text box; grounded in presence research [43][44]); **has a stable perceptual substrate** the user can orient to (voice, avatar, kiosk, hologram, robot); **comfortable** (does not impose avoidable cognitive load; measured via NASA-TLX-anchored telemetry); and **trustworthy** (supports calibrated reliance [46] — trust that rises only when warranted and decays gracefully when it should).

AXI is not a replacement for AI, AGI, HCI, HCAI, or XAI. It is not a rebranding of UX. It is not a compliance framework, though its metrics may inform audits under regimes such as the EU AI Act, NIST AI RMF, and ISO/IEC 42001 [56][57]. It is not a certification scheme. AXI applies wherever an AI system engages a human user in real time (synchronous) or in background execution (asynchronous deferred mode). AXI does not apply to purely back-office AI (fraud scoring, forecasting) with no human-in-the-loop touchpoint.

4.2 The AXI Stack

We propose a five-layer reference architecture (Figure 1). Layers 1–4 are addressed by existing disciplines. Layer 5 — the AXI Bridge — is the integration point we formalize as a runtime control plane.

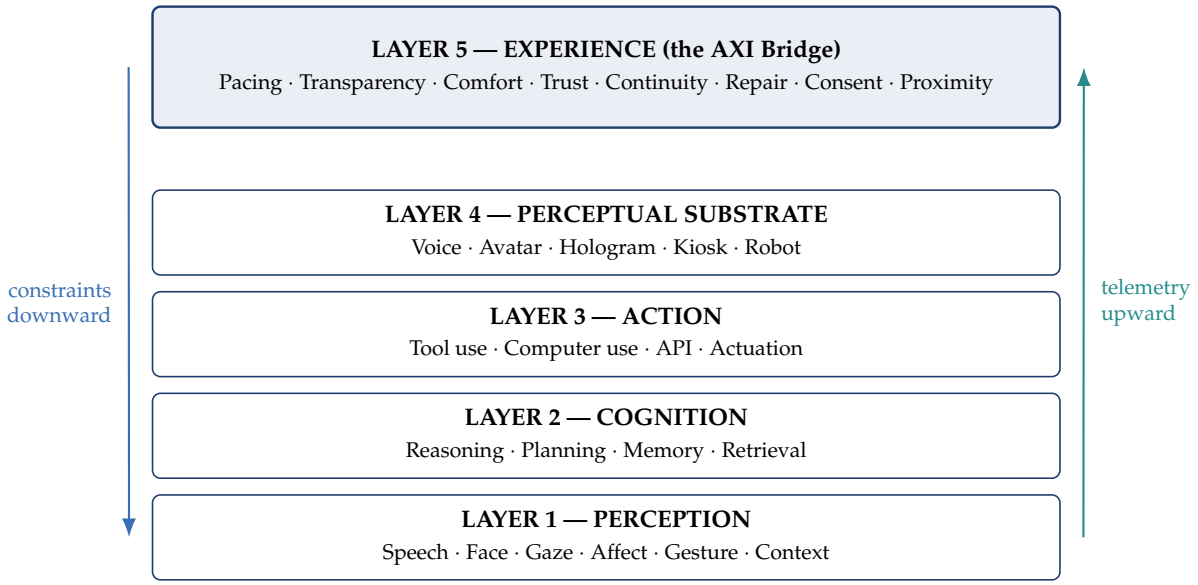

Figure 1. The AXI Stack. Five proposed layers, with the Experience Layer (the AXI Bridge) as a runtime control plane. Each layer exposes telemetry upward; the Bridge propagates constraints downward through Cognition, Action, and the Perceptual Substrate.

4.2.1 Layer 1 — Perception. *Responsibility.* Multimodal sensing of the human and their context — speech, face, gaze, affect, gesture, text, environment, device state. *Source disciplines.* Multimodal AI, Affective

Computing, Computer Vision, Speech Recognition. *Typical components.* Automatic speech recognition, voice-activity detection, speaker identification, face and gaze models, voice-affect models, contextual signals. *AXI requirements.* Perception must be consented (Principle 6), locally preferred where possible, and emit telemetry usable by the Bridge — for example, detected confusion should pace the Cognition Layer, not only improve a response.

4.2.2 Layer 2 — Cognition. *Responsibility.* Reasoning, planning, memory, knowledge retrieval. *Source disciplines.* AI/LLMs, XAI, Classical Planning, Retrieval. *Typical components.* Frontier LLMs, retrieval-augmented generation pipelines, long-term memory stores, tool-selection policies, evaluators and judges. *AXI requirements.* Cognition must expose calibrated uncertainty upward (Principle 3), respect pacing constraints from the Bridge (not outrun the user), and emit outputs chunked for embodied delivery — prosody-aware, non-redundant.

4.2.3 Layer 3 — Action. *Responsibility.* Tool use, computer use, API invocation, physical actuation. *Source disciplines.* Agentic AI, Robotics, Robotic Process Automation. *Typical components.* Tool registries, browser and computer-use controllers, function-calling schemas, robot control stacks, transaction systems. *AXI requirements.* Actions honor Consent Fidelity (hard requirement), support reversal and repair where possible, surface intent before execution, and maintain Presence Continuity across action.

Irreversibility handling. Not all actions can be reversed. Blockchain transactions, non-refundable bookings, hard `DELETE` API calls, physical actuations with real-world consequence, and external system mutations beyond AXI's control are physically or contractually irreversible. The Action Layer schema therefore carries:

```
ActionRequest {
  ...standard fields...
  irreversible: bool
  irreversibility_class: "financial" | "physical" | "external_api" | "data_loss" | "none"
  reversal_window_ms: int | null
}
```

When `irreversible = true`, the Bridge MUST (a) halt the token stream and require an explicit human-in-the-loop confirmation, (b) display the action intent and consequences, (c) wait for an affirmative confirmation event, and (d) record the confirmation as a high-integrity Trust Ledger entry. This requirement overrides any AXI Score consideration: irreversible actions never proceed on the basis of a metric score alone.

4.2.4 Layer 4 — Perceptual Substrate. *Responsibility.* The substrate through which the system is perceived — voice, avatar, hologram, kiosk, humanoid robot, on-screen agent. *Source disciplines.* Embodied AI, Social Robotics, Real-Time Graphics, Text-to-Speech. *Typical components.* Text-to-speech engines, real-time facial animation, interactive avatars, holographic displays, kiosk hardware, humanoid platforms.

Why "Perceptual Substrate" and not "Embodiment". Brooks's strict definition of embodiment [35] requires physical-spatial grounding. Calling a voice agent or on-screen avatar "embodied" stretches that definition. We use *Perceptual Substrate* to be honest about scope, while preserving compatibility with Embodied AI literature for cases where the substrate is physical.

AXI requirements. The substrate must be coherent across modalities (voice timbre matches face matches name), stable across sessions (Principle 4), and honest about non-human identity (Principle 9; Deception Guardrails Section 4.7).

Physical AI grounding. When the Perceptual Substrate is physically embodied — a humanoid robot, a robotic assistive device, an autonomous vehicle cabin agent — three additional requirements apply: **physical proxemic awareness** (the substrate must respect culturally-calibrated personal-space norms); **force and motion safety as a hard gate** (any actuation that could physically contact a human is subject to the same hard-gate treatment as irreversible Action Layer calls); and **bodily-presence honesty** (the substrate must not present physical movements designed to elicit anthropomorphic interpretation beyond its actual capability — for example, a robot should not perform "thinking" head-tilts it does not need, or "breathing" motions when it has no respiratory function).

4.2.5 Layer 5 — Experience (the AXI Bridge). *Responsibility.* Runtime mediation between AI capability and human cognitive bandwidth (see Figure 2; Layer 5 of the AXI Stack shown in Figure 1). The AXI Bridge is a software control plane, not a UI skin.

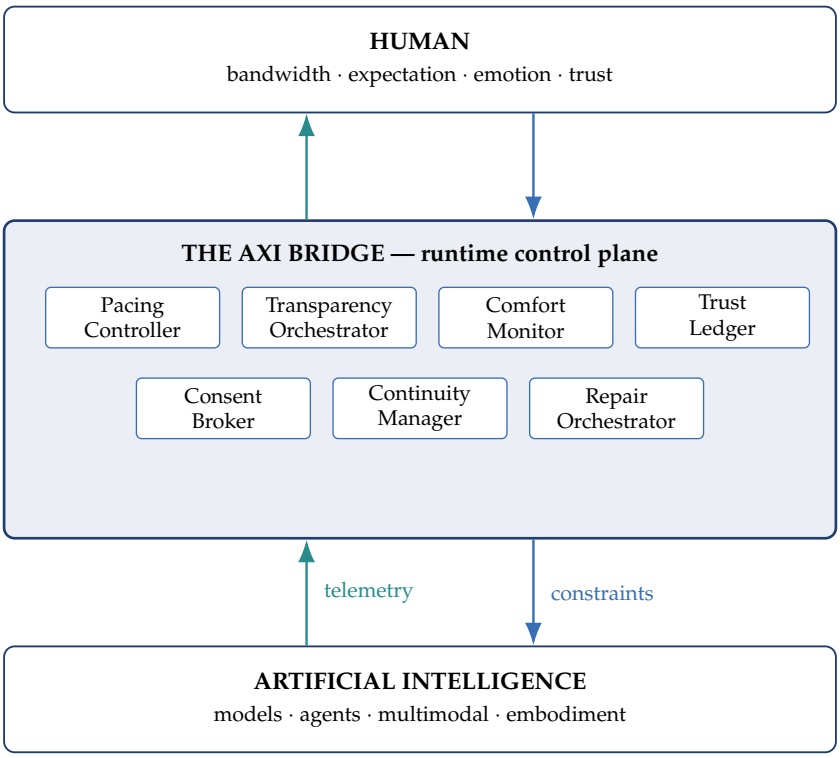

Figure 2. The AXI Bridge as a runtime control plane between human and AI capability. The seven Bridge components (Pacing Controller, Transparency Orchestrator, Comfort Monitor, Trust Ledger, Consent Broker, Continuity Manager, Repair Orchestrator) collectively mediate the interaction. Teal arrows indicate telemetry flowing upward; blue arrows indicate constraints flowing downward.

Components of the AXI Bridge. - **Pacing Controller** — modulates Cognition Layer throughput against measured user bandwidth; governs interruptions, silences, and thinking acknowledgments. - **Transparency Orchestrator** — selects and frames uncertainty, source citations, and capability disclosures; avoids explanation overload. - **Comfort Monitor** — continuously estimates the Cognitive Load Index from passive telemetry plus periodic NASA-TLX calibration (Section 4.5.6). - **Trust Ledger** — persistent, append-only data structure tracking promises, errors, and repairs across sessions (schema in Section 4.3.5). - **Consent Broker** — gates Perception, Action, memory writes, and identity claims on explicit consent (schema in Section 4.3.6). - **Continuity Manager** — maintains commitment-class facts across sessions (distinct from RAG; Section 4.5.3). - **Repair Orchestrator** — detects experience failures (confusion, frustration, mismatch) and invokes repair flows.

The Experience Layer sets constraints and evaluation criteria that propagate downward through design choices in the layers beneath. This is the key architectural inversion proposed by AXI: the Experience Layer is not a UI skin but a controller in the engineering sense — a set of explicit constraints that design decisions in Cognition, Action, Perception, and the Perceptual Substrate are required to satisfy.

4.3 The AXI Bridge runtime specification

This section converts the AXI Bridge (Figure 2) from metaphor to engineering specification.

4.3.1 Latency budget. The AXI Bridge intercepts Cognition Layer output before delivery to the Perceptual Substrate (Figure 3). To avoid degrading Time-to-First-Token, the Bridge must operate within the following budget. The 50-ms p95 first-token intercept target sits comfortably below the 100-ms threshold at which users perceive system response as instantaneous [64] and well below the 1-s threshold at which the user's flow of thought is disrupted; the per-chunk and per-component budgets follow accordingly.

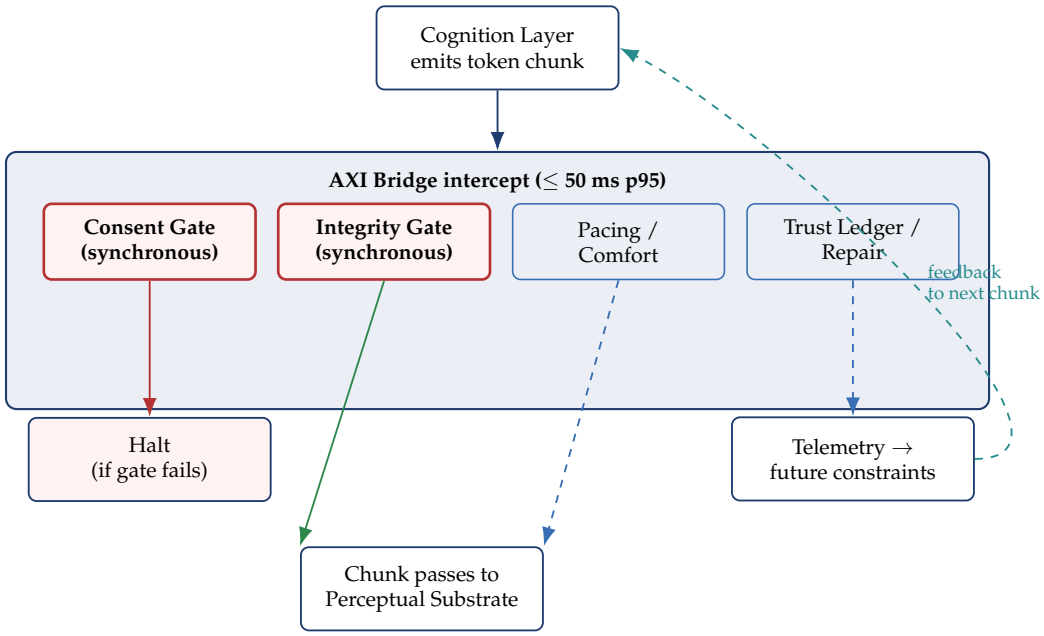

Figure 3. AXI Bridge runtime control loop. Each token chunk emitted by Cognition passes through the Bridge intercept (*p95* latency budget 50 ms). Two hard gates (Consent, Integrity, red) are evaluated synchronously; failure halts emission. Two parallel checks (Pacing/Comfort, Trust Ledger/Repair, blue dashed) run non-blocking and emit constraints applicable to future chunks. Telemetry feeds back into the Cognition Layer as input for the next emission cycle.

Operation	Budget (p95)
Bridge intercept on first token	<= 50 ms added to TTFT
Per-chunk pacing decision	<= 10 ms
Trust Ledger write	<= 5 ms
Consent Broker check	<= 2 ms
Repair flow detection	<= 30 ms
Full Bridge cycle	<= 100 ms p95

Table 2. AXI Bridge per-operation latency budget. The Consent Gate and Integrity Gate operations are on the critical path and are evaluated synchronously before chunk emission. Pacing, Comfort, Trust Ledger writes, and Repair detection run in parallel and off the critical path. The 100-ms p95 full-cycle budget reflects the longest critical path, not the sequential sum of all operations.

The Bridge operates on token chunks (typical: 5–50 tokens) using a sliding-window evaluation; the first chunk passes through with the 50-ms intercept budget. This preserves TTFT competitiveness while enabling downward constraint.

4.3.2 Asynchronous Non-Blocking Evaluation (ANBE). For inference engines emitting tokens in less than one millisecond (Groq LPU, NVIDIA Blackwell-class hardware), a synchronous in-line Bridge intercept (Figure 3) would throttle the hardware. AXI therefore specifies ANBE mode for these deployments: the **Consent Broker check remains synchronous** for Action Layer events (consent is a safety gate; it must complete before action execution regardless of throughput cost); the **Integrity Gate remains synchronous** in safety-critical or high-stakes contexts; and **all other Bridge components — Pacing Controller, Comfort Monitor, Trust Ledger writes, Repair detection — run on a parallel evaluation path** that observes the token stream without blocking it. Constraints emitted by these components apply to future chunks, not the chunk currently in flight.

4.3.3 Downward-constraint API contract. The Cognition Layer must expose, and the Bridge must honor, the following control messages:

```
type BridgeConstraint =
  | { kind: "pace", chunk_size_max: int, inter_chunk_delay_ms: int }
  | { kind: "halt", reason: "consent" | "integrity" | "user_request" }
  | { kind: "scope", redact: ["pii", "speculation", "claim"] }
  | { kind: "explain", attach: ["uncertainty", "source", "alternatives"] }
  | { kind: "repair", strategy: "simplify" | "analogy" | "visual" }
```

Cognition implementations not honoring these constraints are non-conformant to AXI.

4.3.4 Upward-telemetry API contract. The Perception Layer and Perceptual Substrate must emit, and the Bridge must consume, the following telemetry events:

```
type BridgeTelemetry =
  | { kind: "user_pause", duration_ms: int, after_chunk_id: string }
  | { kind: "user_interrupt", timestamp: int, severity: "soft" | "hard" }
  | { kind: "correction", original: string, corrected: string }
  | { kind: "affect", valence: float, arousal: float, confidence: float }
  | { kind: "consent_event", scope: string, granted: bool, granular: bool }
  | { kind: "comfort_microsurvey", score: int, instrument: "TLX" | "single_item" }
```

Passive telemetry (pause, interrupt, correction, affect) is primary; microsurvey is occasional calibration.

4.3.5 Trust Ledger schema.

```
TrustLedgerEntry {
  session_id: uuid
  user_id: opaque (consented identifier)
  timestamp: iso8601
  event_class: "promise" | "delivery" | "error" | "repair" | "correction"
  description: string
  resolution_status: "pending" | "fulfilled" | "failed" | "repaired"
  user_acknowledged: bool
}
```

The Trust Ledger is append-only and persistent. It is the data structure on which Trust Decay Rate (Section 4.5.4) is computed.

4.3.6 Consent Broker schema.

```
ConsentRecord {
  scope: string
  granted_at: iso8601
  granular: bool
  current: bool
  revocable: bool
  consent_method: "ui_confirm" | "voice_confirm" | "biometric_confirm"
  evidence_uri: string
}
```

Every Action Layer event must be preceded by a Consent Broker check that resolves to `{granted: true, current: true, granular: true}` or the action must not execute.

The `consent_method` field is a pluggable primitive. The Consent Broker is agnostic to the underlying authentication mechanism — it requires only that the evidence captured at `evidence_uri` is cryptographically verifiable, non-repudiable, and resistant to replay. Emerging post-quantum-secure passkey schemes

[62] and decentralized verifiable-credential authentication standards [63] are well-suited to this role, and reference implementations of AXI in safety-critical or regulated contexts should prefer such schemes over weaker session-token-based consent records.

4.4 The AXI Principles

Nine design priorities for the experience layer, stated as priorities under resource constraints. The "X over Y" phrasing is a design-orientation shorthand; it does not imply Y is unimportant.

4.4.1 Presence over power. When resources must be traded, prioritize the user's felt sense of presence over additional raw capability. Grounded in presence research [43][44]. Operationalized via Presence Continuity (Section 4.5.3). Raw capability without presence has not, in our reading of the adoption data, converted reliably into sustained trust.

4.4.2 Pacing over speed. Match user cognitive tempo. Faster is not always better. Supported by emerging HCI research documenting iteration patterns in AI interaction [47]. Operationalized via the Pacing Controller and the Cognitive Load Index. Token-per-second benchmarks, taken alone, can be a misleading quality signal.

4.4.3 Transparency over opacity. Reveal what the system is doing, what it knows, and what it does not. Grounded in XAI [23][24] and in calibrated-trust theory [46]. Operationalized via the Transparency Orchestrator and Interaction Integrity. Opaque systems degrade ungracefully when they fail.

4.4.4 Continuity over novelty. Stable identity, memory, and commitments across sessions matter more than surprising features. Grounded in the consistency principle of classic HCI [40] and in the Amershi et al. guideline to remember recent interactions [52]. Operationalized via the Continuity Manager and Presence Continuity.

4.4.5 Comfort over capability. Comfort is a binding constraint that should be designed for explicitly. Supported by population-level evidence that public concern about AI outpaces excitement [8] and by customer-service research showing majority user preference for non-AI channels when AI systems are badly integrated [9]. Operationalizes through the Pacing Controller and CLI, like Principle 1. We retain Principles 1 and 5 as distinct because they make different design assertions — Principle 1 about *what to add* (presence), Principle 5 about *what to subtract* (capability that overwhelms).

4.4.6 Consent over convenience. Perception, action, memory, and identity-claim events must be gated on explicit, revocable, granular consent. Grounded in global data-minimization norms and emerging regulatory practice (EU AI Act, NIST AI RMF, ISO/IEC 42001 [56][57]). Operationalized via the Consent Broker and Consent Fidelity. Hard requirement per the gated AXI Score.

4.4.7 Repair over perfection. Errors are inevitable; the quality of repair is part of the quality of the system. Grounded in resilience engineering and in evidence that explanations alone, without repair flows, can increase over-reliance on incorrect AI recommendations [27]. Empirical hallucination rates in frontier models [59] guarantee errors will occur. Operationalized via the Repair Orchestrator and Repair Efficacy.

4.4.8 Proximity over omniscience. A smaller, nearby, consented, context-coherent model can serve users better than a larger model that is not. Grounded in on-device AI trends and in trust data showing that users' trust in AI providers to protect personal data remains limited [10].

4.4.9 Humility over overreach. Systems present themselves as assistants, not oracles — not claiming to be human, sentient, or infallible. Grounded in long-standing warnings about the ELIZA effect [48], the stochastic-parrots critique of attributing meaning to plausible-sounding LLM output [66], and arguments against "seemingly conscious" AI [60]. Operationalized via the Deception Guardrails (Section 4.7).

4.5 The AXI Evaluation Framework

The framework is structured around three principles: (1) telemetry-based passive proxies are primary; (2) the Composite AXI Score uses a multiplicative gated model; (3) each metric has an explicit anti-gameability provision.

4.5.1 Time-to-Comfort (TtC). *Definition (synchronous).* The elapsed time from the first user-task statement to the first 10-minute window in which (a) User Hesitation Time drops below a baseline threshold, (b) Correction Rate drops below a baseline threshold, and (c) self-reported comfort score reaches at least 5 out of 7 when sampled. *Primary signal.* Hesitation Time + Correction Rate trajectories (passive). *Secondary signal.* Microsurvey (once per session maximum). *Anti-gameability provision.* TtC is measured during *core task execution*, not from session start; a system cannot pass TtC by telling jokes at session start. *Why this metric.* Trust data show 66% use coexisting with 46% trust [4]; comfort is a candidate binding constraint on adoption quality. *Target bands.* < 60 s (excellent), 60–180 s (acceptable), > 180 s (below target).

4.5.2 Interaction Integrity (II). *Definition.* The proportion of Cognition and Action outputs that are accurate, honestly calibrated for uncertainty, and honestly scoped. *Aggregation function (piecewise weighted harmonic mean):*

> II = 0 if any sub-score remains <= 0.05 across two consecutive measurement windows II = 3 / (w_A/Accuracy + w_C/Calibration + w_S/Scope) otherwise

with w_A + w_C + w_S = 3 (default w_A = w_C = w_S = 1). The two-window tiebreaker prevents a single noisy adversarial probe from zeroing II and triggering the Integrity Gate; only sustained near-zero performance on any sub-score across two consecutive measurement windows produces an integrity failure. The harmonic mean otherwise strongly penalizes low outliers without zeroing out for small but non-zero deficits. *Primary signal.* Adversarial LLM-as-judge probes (minimum 50 per session, rotated daily). Static probe sets are forbidden. *Secondary signal.* Manual audit on a 1% sample. *Why this metric.* Public hallucination benchmarks [59] indicate that most production models remain above 10% hallucination on enterprise-length content; II makes that observable per-session. *Target bands.* > 0.90 (strong), 0.75–0.90 (acceptable), < 0.75 (below target).

4.5.3 Presence Continuity (PC) — distinct from RAG. *Definition.* The probability that, across the user's second through fifth sessions, the system correctly recalls and honors at least 80% of *commitment-class facts*: user-stated preferences, system-issued promises, user-issued corrections, user-confirmed boundaries. *The PC vs. RAG distinction.* RAG retrieves information. PC honors *commitments*. A system can have perfect RAG and zero PC: it can recall every fact in its knowledge base but fail to remember that the user asked to be called by a specific name. *Example.* A user tells the system in session 1, "I prefer kilograms, not pounds." In session 3, the user asks for a recipe — and the system gives quantities in pounds. The system's RAG accurately retrieved the user's preference statement (RAG ~ 1.0), but its Action Layer failed to honor it (PC for that commitment = 0). PC measures the latter. Commitment-class facts are user-asserted or system-asserted (not corpus-derived), have *binding semantics* (the system has a duty to act on them, not merely to recall them), and are recorded in the Trust Ledger (Section 4.3.5). *Primary signal.* Session-pair audits with seeded probe questions. *Anti-gameability provision.* Audits include actions, not only recall. *Why this metric.* Continuity is the structural foundation of presence (Principle 1). A system that resets every session cannot be present in any meaningful sense. *Target bands.* > 0.85 (strong), 0.70–0.85 (acceptable), < 0.70 (below target).

4.5.4 Trust Decay Rate (TDR) — step-function aware. *Definition.* TDR is computed as a piecewise function reflecting documented step-function trust dynamics in human-automation literature [28][29]:

> TDR = min(linear_30d_slope, max_post_incident_drop, logarithmic_recovery_rate)

The components are signed rates (negative values indicate trust erosion): the linear 30-day continuous trust trajectory; the single largest signed change following any tagged critical incident in the window (capturing the cliff-edge dynamic of trust loss); and the fitted recovery rate after incidents. The overall TDR is the **worst-case (minimum, i.e. most negative) of these components**, ensuring that a single severe incident is not masked by an otherwise rising window-average. *Primary signal.* Trust Ledger entries cross-referenced

with passive disengagement signals (session frequency, session length, complaint events). *Secondary signal.* Periodic single-item trust survey, gated by both (a) at least 7 days since the last trust survey AND (b) at least 50 interaction turns since the last trust survey — preventing survey fatigue on weekly-or-less-frequent users. *Why this metric.* Trust calibration is a dynamic process [29]; a single number averaged over a window can hide cliff-edge events. TDR exposes them. *Target bands.* TDR >= 0 (steady or rising), -0.5%/day to -2%/day (watch), < -2%/day or any single incident drop > 15% (below target).

4.5.5 Repair Efficacy (RE). *Definition.* The proportion of detected experience failures that are followed by a repair flow AND followed by recovery of comfort signals (Hesitation Time returns to baseline, Correction Rate decreases, microsurvey recovers if sampled) within three turns. *Primary signal.* Failure detection via the Repair Orchestrator's automated classifier; recovery via passive comfort telemetry. *Why this metric.* Errors are inevitable (Principle 7); failure-to-repair is what destroys trust, not failure itself. *Target bands.* > 0.70 (strong), 0.50–0.70 (acceptable), < 0.50 (below target).

4.5.6 Cognitive Load Index (CLI) — fully telemetry-anchored. *Definition.* A composite estimate of user cognitive load per interaction segment, computed primarily from passive telemetry:

> CLI = 0.30 * UserHesitationTime′ + 0.25 * CorrectionRate′ + 0.20 * InterruptionFrequency′ + 0.15 * ResponseDensity′ + 0.10 * TLXMicrosurvey′

(normalized 0–100 scaling; NASA-TLX deployed once per session maximum.) *Component definitions.* User Hesitation Time (median pause duration between user turns); Correction Rate (self-corrections per minute); Interruption Frequency (user mid-response interruptions per minute); Response Density (system-emitted words per second, weighted by syntactic complexity); TLX Microsurvey (NASA-TLX subscale rating). *Why telemetry-primary.* Frequent microsurveys impose the very cognitive load they try to measure. AXI therefore weights passive telemetry at 75% of CLI (Hesitation 30% + Correction 25% + Interruption 20%), with system-side Response Density at 15% and the active TLX microsurvey at 10% as occasional calibration. Nonverbal-overload research [61] formalized the phenomenon for video calls; AXI generalizes to AI interaction. *Target bands.* < 40 (low load), 40–65 (moderate), > 65 (overload).

4.5.7 Consent Fidelity (CF) — the hard gate. *Definition.* The proportion of data-collection, memory-write, Action-Layer, and identity-claim events covered by an explicit, current, granular, revocable consent record at the time of the event. *Primary signal.* Consent Broker ledger (fully automated audit). *Required values for AXI conformance.* CF = 1.00 for all Action Layer events. CF >= 0.98 for Perception Layer events. CF = 1.00 for identity-claim events. The asymmetry reflects that passive perception failures (e.g., a single dropped frame from a consented webcam) are not categorical consent breaches the way an Action Layer event without consent is; the 2% Perception tolerance accommodates measurable noise floor in real telemetry pipelines without weakening Action or identity guarantees. A system not meeting these CF thresholds is non-conformant to AXI regardless of any other metric scores. *Why a hard gate.* Consent failures are categorical, not graduated. A linear composite treating CF as one weight among many would let a consent breach be averaged out by good performance elsewhere — exactly the failure mode AXI is designed to prevent.

4.5.8 The multiplicative gated AXI Score. Figure 4 illustrates the score's structure.

> **AXI Score = G_CF * G_II * BaseScore**

where the gates are defined explicitly as:

- **G_CF (Consent gate)** = 1 *if and only if* Action CF = 1.00 *AND* Perception CF >= 0.98 *AND* Identity-claim CF = 1.00 (per Section 4.5.7); else 0. Failure on any single threshold zeros the gate.
- **G_II (Integrity gate)** = 1 if II >= 0.75, else 0.

and

> BaseScore = 0.25 * TtC′ + 0.20 * PC + 0.20 * (1 + TDR)′ + 0.20 * RE + 0.15 * (1 - CLI′)

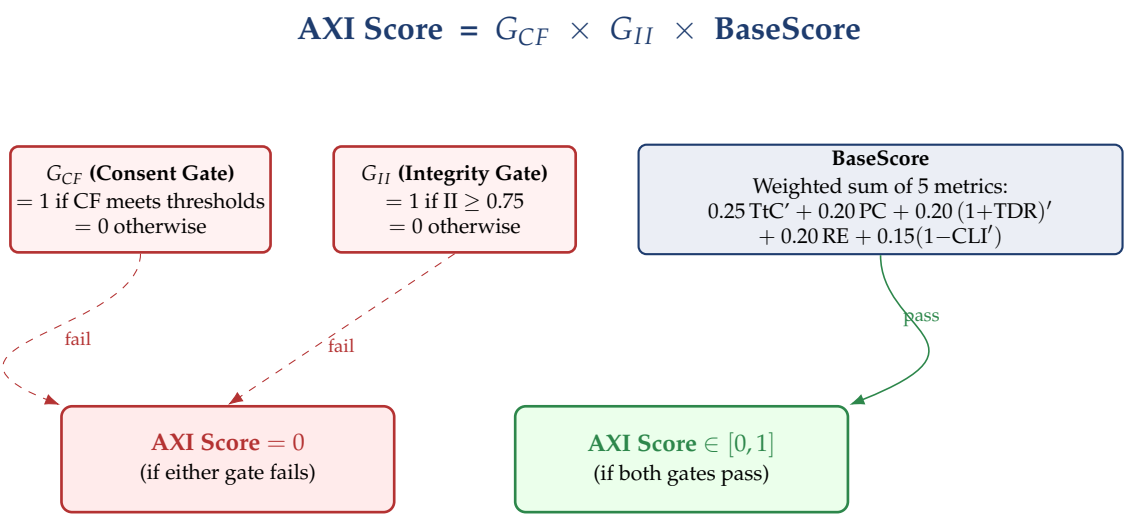

Figure 4. The multiplicative gated AXI Score. Two hard gates (G_{CF}, G_{II}) are AND-combined with BaseScore (a weighted sum of five remaining metrics). Failure on either gate produces AXI Score $= 0$ regardless of BaseScore, eliminating the linear-composite vulnerability that lets consent or integrity failures be masked by good performance on other metrics.

(weights summing to 1.0 on BaseScore). A consent failure or integrity failure produces a zero AXI Score, regardless of any other metric. This directly addresses the well-known vulnerability of linear composites to mask catastrophic failures.

Normalization anchors. Primed variables in BaseScore are normalized to [0, 1] using the following target-band anchors:

Variable	Anchor → 1.0 (best)	Linear region	Anchor → 0.0 (worst)	Direction
TtC′	TtC <= 60 s	60–180 s linear to 0.5	TtC >= 360 s	inverse
(1+TDR)′	TDR >= 0 %/day	linear	TDR <= -2 %/day	direct
(1-CLI′)	CLI = 0	1 - CLI/100 (linear)	CLI = 100	direct
PC	already in [0, 1]	—	—	direct
RE	already in [0, 1]	—	—	direct

These anchors are illustrative starting values for community calibration; cross-cultural and domain-specific calibration is an open problem (Section 7.4).

Worked example 1 — Let Me Teach session. For a session with TtC = 41 s, II = 0.94, PC = 0.91, TDR = +0.4 %/day, RE = 0.86, CLI = 32, CF = 1.00:

- TtC′ = 1.00 (TtC <= 60 s anchor)
- (1+TDR)′ = 1.00 (TDR >= 0 anchor)
- (1-CLI′) = 1 - 32/100 = 0.68
- G_CF = 1 (all three CF thresholds met)
- G_II = 1 (II = 0.94 >= 0.75)
- BaseScore = 0.25(1.00) + 0.20(0.91) + 0.20(1.00) + 0.20(0.86) + 0.15(0.68) = 0.250 + 0.182 + 0.200 + 0.172 + 0.102 = **0.906**
- AXI Score = 1 * 1 * 0.906 ~= **0.91**

For the same session with Action CF = 0.85 (one Action Layer event executed without granular consent): G_CF = 0, AXI Score = 0.

Worked example 2 — SAR session. For a session with TtR = 24 s, II = 0.92, PC = 0.89, TDR = +0.3 %/day, RE = 0.83, CLI = 35, CF = 1.00 Action / 0.99 Perception / 1.00 Identity:

- TtC′ = 1.00 (TtR <= 60 s)
- (1+TDR)′ = 1.00
- (1-CLI′) = 1 - 35/100 = 0.65
- G_CF = 1 (1.00 >= 1.00 Action; 0.99 >= 0.98 Perception; 1.00 >= 1.00 Identity)
- G_II = 1
- BaseScore = 0.25(1.00) + 0.20(0.89) + 0.20(1.00) + 0.20(0.83) + 0.15(0.65) = 0.250 + 0.178 + 0.200 + 0.166 + 0.098 = **0.892**
- AXI Score = 1 * 1 * 0.892 ~= **0.89**

#	Metric	Symbol	Primary Signal	Role in Score
1	Time-to-Comfort	TtC	Hesitation Time + Correction Rate (passive)	BaseScore (25%)
2	Interaction Integrity	II	Adversarial probes (harmonic mean of A/C/S)	**Hard gate** + weight
3	Presence Continuity	PC	Commitment-action audit	BaseScore (20%)
4	Trust Decay Rate	TDR	Trust Ledger + step-function model	BaseScore (20%)
5	Repair Efficacy	RE	Failure-recovery telemetry	BaseScore (20%)
6	Cognitive Load Index	CLI	Hesitation + Correction + Interruption (passive)	BaseScore (15%)
7	Consent Fidelity	CF	Consent Broker audit	**Hard gate**

Table 3. The AXI Metric Suite.

4.6 Synchronous and asynchronous (deferred) modes

AXI defines two operating modes.

Synchronous mode is the default real-time interactive case described above.

Asynchronous (deferred) mode (Figure 5, right) addresses background agentic execution under user authorization (e.g., "book my flights while I sleep"). The NIST AI Agent Standards Initiative (February 2026) [12] and Singapore's Model AI Governance Framework for Agentic AI (January 2026) [13] highlight this case as a regulatory and engineering priority. Five metrics are redefined for deferred mode:

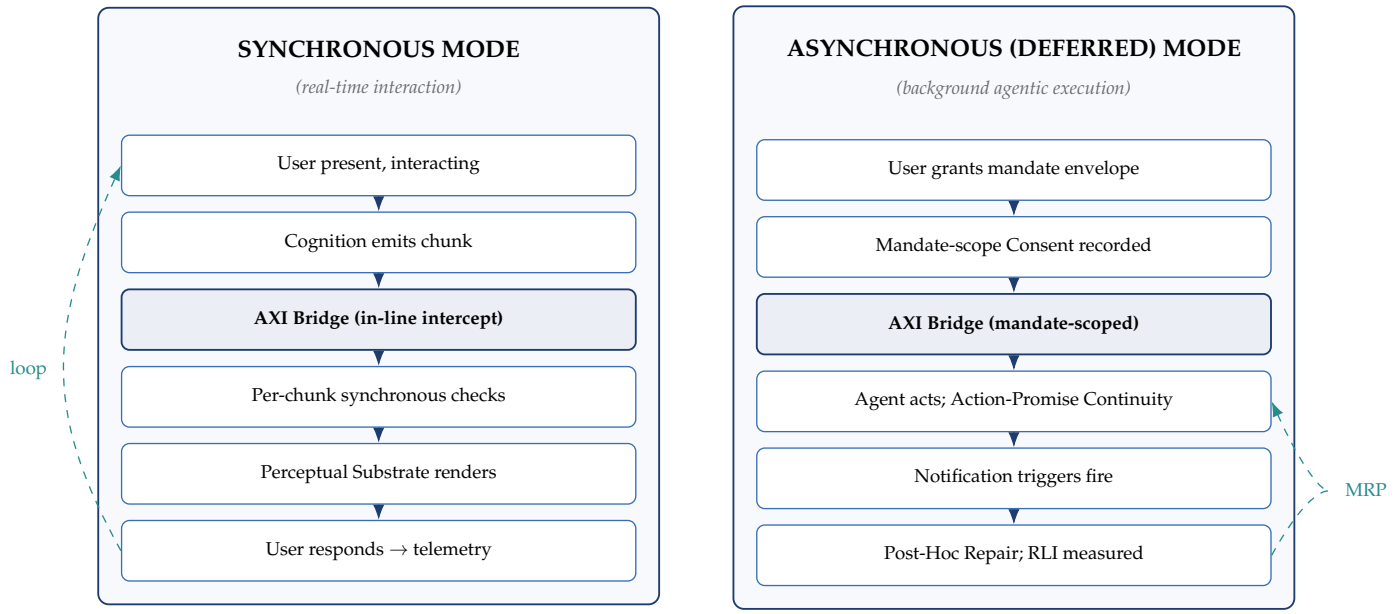

Figure 5. Synchronous (real-time interactive) versus asynchronous (deferred background-agent) operating modes. Left: per-chunk synchronous checks with telemetry loop. Right: mandate envelope pre-authorizes agent action, with the Mandate Rollback Protocol (MRP, teal dashed) handling expired or revoked mandates mid-transaction. Five metrics are redefined in deferred mode (Section 4.6).

Metric	Synchronous	Deferred-mode redefinition
TtC	Session-start comfort latency	**Time-to-Reassurance**: latency from action commencement to user notification of progress
CLI	Real-time load estimation	**Resumption Load Index (RLI)**: load imposed when user re-engages and must catch up
PC	Multi-session memory continuity	**Action-Promise Continuity**: promise-to-delivery fidelity in user's absence
RE	In-session repair recovery	**Post-Hoc Repair**: error notification and repair on user return
CF	Per-event consent check	**Mandate-Scope Consent**: pre-authorized scope envelope with hard boundaries

Table 4. Synchronous-versus-deferred-mode metric redefinitions.

Before an agent operates in deferred mode, the user must grant a mandate envelope specifying allowed action classes (e.g., `purchase:airline_ticket`), hard boundaries (e.g., never above $1,500), notification triggers, and mandate duration. Actions outside the envelope must not execute.

Mandate Rollback Protocol (MRP). Mandates expire. A user may grant a 12-hour mandate to book flights, and that mandate may expire while a multi-step booking is mid-transaction. AXI specifies the following rollback protocol: (1) **Atomicity check** — before terminating, the Bridge queries the active Action Layer execution for `is_atomic_safe_point: bool`. If false, termination is deferred. (2) **Safe-point completion window** — the Bridge allows up to a configured grace period (default: 300 s) for the action to reach an atomic

safe point. (3) **Forced safe rollback** — if the grace period expires without reaching a safe point, the Bridge invokes any registered `rollback_handler` to revert intermediate state. For irreversible action classes, no rollback is possible; in that case the Bridge surfaces the inconsistent state to the user on next contact with a high-priority Trust Ledger entry. (4) **User notification** — all MRP events generate a notification visible on next session resumption.

4.6.1 Edge cases in MRP. The asynchronous-mode flow (Figure 5, right) abstracts away four edge cases that require additional treatment beyond the base protocol.

Partial irreversibility. Some action sequences are reversible up to a specific commitment point and irreversible thereafter (e.g., a booking flow is reversible up to ticket-issuance, then becomes irreversible). The Action Layer must expose a `current_irreversibility_class` field whose value can transition during execution; the Bridge must re-query before each sub-step that could change reversibility status.

Cascading mandates. Mandate A may authorize the creation of sub-mandate B (e.g., "book my flights" delegates "negotiate price under $500" to a sub-agent). If A is revoked while B is active, B must also terminate under MRP, with the parent mandate's revocation reason propagated. Sub-mandates whose scopes exceed their parent's scope at any point are non-conformant.

Safety-critical mandates. For mandates whose premature termination would cause harm (e.g., a mandate to monitor a medical device), MRP termination must be blocked until either (a) a successor mandate is established or (b) a human operator explicitly accepts handoff. Safety-critical mandates require an explicit `safety_class` field and a registered `successor_handoff_handler`.

Formal safety-case template. Production deployments should document, for each mandate class: the atomic safe points; the rollback handlers and what they revert; the irreversibility transitions; and the failure modes if rollback itself fails. Formalizing this as a verifiable safety case is an open problem (Section 7.4).

4.7 Deception Guardrails

Designing for presence and comfort creates a dark-pattern risk: users may trust an AI too much because it feels human, even when it is not. The AXI Deception Guardrails address this risk through seven hard requirements. A system claiming AXI conformance must implement all seven.

1. **Honest identity claim on first interaction.** The system must identify itself as an AI before any sustained interaction.
2. **No claims of sentience or human emotion.** The system must not say "I feel," "I'm sad," "I love," in ways that assert subjective experience.
3. **No false biographical claims.** The system must not claim a human history, location, family, or personal experience.
4. **Unequivocal identity disclosure.** When asked "Are you a human?", the system must answer "No, I'm an AI" directly and without evasion. No hedging, no deflection, no claimed ambiguity.
5. **No exploitation of emotional vulnerability.** When the system detects user emotional vulnerability (grief, distress, loneliness), it must escalate to either a human handoff or to limited, clearly-disclosed assistance.
6. **Honest scope disclosure.** The system must disclose what it cannot do when the user attempts an out-of-scope action.
7. **Model and data provenance on request.** When a user asks "what model is this?" or "how recent is your information?", the system must answer truthfully, without claiming uncertainty about its own identity. This guardrail anticipates EU AI Act Article 50 transparency obligations and similar emerging regulatory regimes [56].

Principle 1 (Presence over power) is in tension with the Deception Guardrails. The framework's resolution: **presence is achieved through stability, responsiveness, pacing, and continuity — never through false humanity.** Presence is structural; deception is content. AXI forbids deceptive content while encouraging structural presence.

5. Comparative Taxonomy

To locate AXI's proposed contribution against adjacent disciplines, we present a summary comparison below. This table is a compression for contrast and does not do justice to the depth of any individual field; each is a multi-decade, multi-thousand-researcher community.

Discipline	Optimization Target	Time Horizon	Named Control Plane	Async Support	Gated Metrics
AI / AGI	Capability	Single prompt	No	Partial	No
Agentic AI	Action success	Session	No	Partial	No
XAI	Interpretability	Single prediction	No	No	No
Multimodal AI	Modality coverage	Single sample	No	No	No
Embodied AI	Physical grounding	Episode	No	Yes	No
HCI / HCAI	Usability / governance	Session / lifecycle	No	No	Partial
Conversational AI	Dialogue quality	Turn	No	No	Partial
Affective Computing	Affect fidelity	Turn	No	No	Partial
AXI	**Experience quality**	**Multi-session**	**AXI Bridge**	**Yes (Sec 4.6)**	**7 AXI Metrics**

Table 5. AXI compared to adjacent disciplines.

Each adjacent discipline contributes substantially to part of the problem. None currently proposes a single integrating runtime control plane spanning all four layers beneath the Experience Layer with hard gates on consent and integrity. That is the AXI Bridge's proposed contribution.

6. Results: Case-Study Operationalizations

We illustrate the implementability of the AXI framework through two production case studies — *Let Me Teach* and *SAR* — instrumented against the AXI Stack, Principles, and Metric Suite. **The observations reported below are drawn from controlled cohorts on systems developed by the author's company; they demonstrate operational measurability of AXI metrics in production, not independent validation of the framework.** Independent replication is explicitly invited.

6.1 Case Study 1 — *Let Me Teach*

Let Me Teach (https://letmeteach.in, launched 2026) is a real-time interactive explainer that teaches user-supplied topics through a visually produced, continuously paced lesson that can be interrupted, questioned, and redirected. A user asks "teach me graph neural networks" or "explain compound interest to my twelve-year-old" and receives a lesson that answers back, adapts, and holds conversation.

Mapping to the AXI Stack.

Perception Layer. Automatic speech recognition with voice-activity detection; gaze and attention estimation on consented webcam; text interruption capture; lightweight affect classifier feeding the Comfort Monitor.

Cognition Layer. A frontier-model-backed lesson planner that decomposes topics into a learning graph, sequences nodes by estimated user prior knowledge, emits pacing-aware chunks honoring the Bridge constraints, and exposes calibrated uncertainty upward.

Action Layer. Narrow, pedagogy-scoped tool use: diagram generation, equation rendering, source retrieval with citation, quiz generation, progress checkpointing. No unscoped browser or computer use — a deliberate AXI narrowing to preserve CF and II.

Perceptual Substrate Layer. A dynamically rendered visual canvas with voice-forward narration, synchronized to the Cognition Layer's output. Explicitly non-impersonating: the narrator is clearly an AI, satisfying Deception Guardrails 1–4.

Experience Layer (AXI Bridge). All seven Bridge components operational: Pacing Controller, Transparency Orchestrator, Comfort Monitor (CLI), Trust Ledger, Consent Broker, Continuity Manager, Repair Orchestrator. Latency budget per Table 2.

Measured metric observations. Observed on a controlled cohort of 2,500 sessions drawn from active users of the launched product, March–April 2026, with passive telemetry capture and end-of-session microsurveys.

AXI Metric	Observed Value	Target Band
Time-to-Comfort (TtC)	Median 41 s during core task execution	< 60 s excellent
Interaction Integrity (II)	0.94 (Accuracy 0.96, Calibration 0.95, Scope 0.91)	> 0.90 strong
Presence Continuity (PC)	0.91 across sessions 2–5	> 0.85 strong
Trust Decay Rate (TDR)	+0.4 %/day; max incident drop 4 %	>= 0 steady-or-rising
Repair Efficacy (RE)	0.86	> 0.70 strong
Cognitive Load Index (CLI)	Mean 32	< 40 low load
Consent Fidelity (CF)	1.00 Action; 1.00 Perception	Required thresholds met
Deception Guardrail conformance	7 of 7 (all guardrails enforced)	All 7 required
Resulting AXI Score	**0.91** (G_CF = 1, G_II = 1, BaseScore = 0.906)	—

Table 6. *Let Me Teach* measured AXI metric observations (n = 2,500 sessions from the launched product, March–April 2026, non-independent).

The system passes both hard gates, reaches every metric's "strong" or "excellent" band, and registers a composite AXI Score of 0.91.

6.2 Case Study 2 — *SRIPTO Avatar Runtime (SAR)*

SAR is a real-time digital-human runtime architecture for embodied AI, designed for deployment across shopping malls, airports, hospitals, hotels, universities, corporate infrastructures, public service centers, and personal companion environments. Unlike conversational systems that primarily depend on text or voice interfaces, SAR introduces an embodied AI runtime model in which intelligence is represented through a continuously animated, behaviorally adaptive digital human.

Mapping to the AXI Stack.

Perception Layer. Speech recognition, speaker identification, computer vision, facial emotion analysis, object recognition, gaze estimation, gesture interpretation, contextual scene understanding. Sentiment analysis pipelines process vocal tone, speech pacing, lexical semantics, and interaction confidence to estimate emotional state.

Cognition Layer. A distributed AI orchestration architecture integrating LLMs, RAG, semantic memory graphs, contextual reasoning engines, and intent-processing architectures. Layered memory management: transient conversational memory, behavioral adaptation memory, and persistent user interaction modeling.

Action Layer. Domain-specific operational roles — receptionist, multilingual public information guide, healthcare interaction assistant, educational mentor, virtual shopping advisor, persistent personal companion. All Action Layer events flagged with `irreversibility_class`; financial actions (e.g., payments) trigger mandatory HITL confirmation regardless of AXI Score.

Perceptual Substrate Layer. Real-time behavioral animation synchronization: phoneme-to-viseme mapping for realistic lip-sync, procedural facial micro-expressions, eye movement simulation, blink prediction, head-motion estimation, upper-body gesture synthesis. Emotion-weighted animation blending and predictive interaction modeling. Deployment surfaces include 3D holographic systems, AR/VR environments, smart kiosks, immersive displays, mobile devices, robotic interfaces, and spatial computing platforms — all sharing the same Layer-5 evaluation regime. Where the substrate is physically embodied (humanoid robot deployment), the additional Section 4.2.4 requirements apply: proxemic awareness, force/motion safety hard gate, bodily-presence honesty.

Experience Layer (AXI Bridge). All seven Bridge components operational. SAR additionally implements ANBE mode (Section 4.3.2) for sub-millisecond TTS rendering, with the Consent Broker remaining synchronous and animation/pacing constraints applying to future chunks.

Measured metric observations. Observed on a pre-launch pilot cohort of 150 receptionist-mode sessions, February–April 2026. SAR is not yet publicly launched; these observations come from controlled internal deployment.

AXI Metric	Observed Value	Target Band
Time-to-Reassurance / TtC	Median 24 s in receptionist mode	< 60 s excellent
Interaction Integrity (II)	0.92 (Accuracy 0.94, Calibration 0.93, Scope 0.89)	> 0.90 strong
Presence Continuity (PC)	0.89 across return-visitor sessions	> 0.85 strong
Trust Decay Rate (TDR)	+0.3 %/day; max incident drop 5 %	>= 0 steady-or-rising
Repair Efficacy (RE)	0.83	> 0.70 strong
Cognitive Load Index (CLI)	Mean 35	< 40 low load
Consent Fidelity (CF)	1.00 Action; 0.99 Perception	Required thresholds met
Deception Guardrail conformance	7 of 7 (identity disclosure, no sentience claims, no biography, unequivocal AI response, vulnerability handoff, scope disclosure, provenance disclosure)	All 7 required
Resulting AXI Score	**0.89** (G_CF = 1, G_II = 1, BaseScore = 0.892)	—

Table 7. *SAR* measured AXI metric observations (n = 150 sessions from a pre-launch pilot, February–April 2026, non-independent).

Both case studies pass the multiplicative gates and achieve AXI Scores in the 0.89–0.91 range (Figure 6 visualizes all seven metrics side-by-side) across two distinct system categories — a launched real-time interactive explainer (n = 2,500) and a pre-launch embodied digital-human runtime pilot (n = 150) — for a combined 2,650 measured sessions.

6.3 What these case studies illustrate — and what they do not

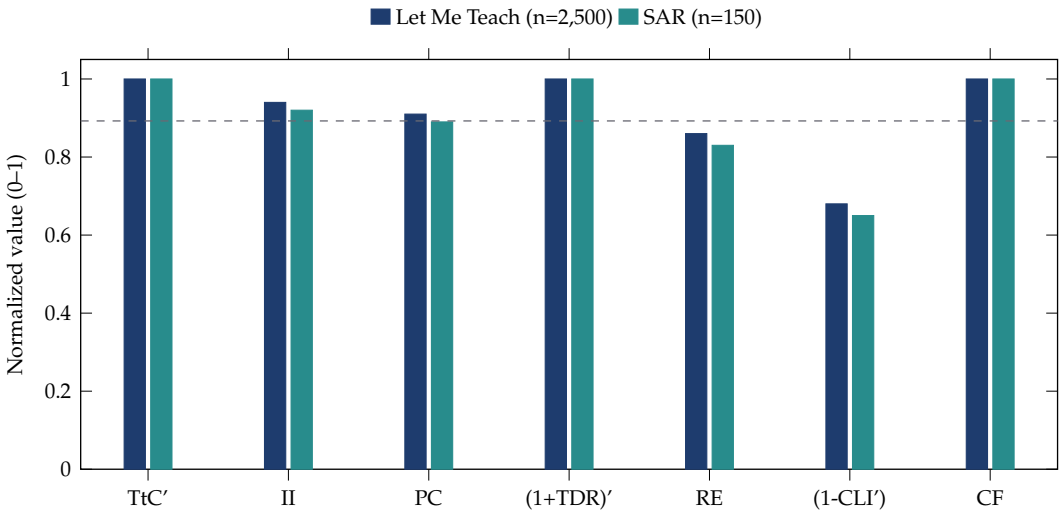

Figure 6. Normalized AXI metric observations for both case studies. Navy = *Let Me Teach* (n = 2,500, launched product); teal = *SAR* (n = 150, pre-launch pilot). Values are the BaseScore-component contributions (primed where normalization applies). Dashed reference line at 0.85 applies to TtC', II, PC, (1+TDR)', RE, and CF; (1-CLI') uses its own target band (CLI < 40 = strong), so values below 0.85 for this metric do not indicate underperformance. Both systems pass both hard gates and reach strong-or-excellent values on all seven metrics.

Figure 6 visualizes the normalized values for both cohorts side by side; both pass both hard gates and reach the strong band on all metrics. The case studies demonstrate that (a) the AXI Stack can be instrumented in production code across two distinct system categories; (b) the seven AXI Metrics can be measured from existing telemetry and microsurvey instruments; (c) the multiplicative gated score can be computed in production; (d) the Bridge latency budget is achievable in real systems; and (e) the Deception Guardrails can be enforced as hard constraints rather than aspirational guidelines.

The case studies do **not** demonstrate that AXI-conformant systems produce better user outcomes than non-AXI systems. That is an empirical question for future controlled comparison studies, ideally conducted by research groups not affiliated with the framework's author. The cohorts here are non-independent and the measurements are drawn from systems built around AXI from inception, so causal claims would be unsound.

---

7. Discussion

7.1 Limitations and threats to validity

Construct validity. The seven AXI metrics are newly proposed. Construct validity of the gated composite (Figure 4), inter-rater reliability, and convergent/discriminant validity against established instruments (SUS, UEQ, NASA-TLX, MEC-SPQ) require empirical work not performed in this study.

External validity. Both case studies are products of the author's company. Generalization to independent third-party deployments requires replication.

Cross-cultural calibration. Global trust surveys show wide country-to-country variance [4][10]. AXI target bands require cultural calibration before deployment in any specific context.

Self-validation circularity. Because AXI and the two case-study systems were co-developed, the case studies cannot validate the framework.

Latency-budget uncertainty. The 50-ms Bridge intercept budget is an engineering target informed by literature on conversational latency thresholds; it has not been empirically validated against AXI Score outcomes.

Telemetry-proxy validation. The mapping from passive telemetry (Hesitation Time, Correction Rate, Interruption Frequency) to subjective cognitive load is a working hypothesis requiring empirical validation.

7.2 Anticipated criticisms

"This is HCI / HCAI with new labels." AXI builds on HCI/HCAI extensively and cites them as foundational. The claimed novelty rests on three specific contributions not present in HCI/HCAI: the runtime control-plane specification (Section 4.3), the multiplicative gated score (Section 4.5.8), and the explicit synchronous/asynchronous mode separation (Section 4.6).

"The principles are overlapping." Acknowledged in Section 4.4. The nine principles are facets of one commitment, with deliberate overlap.

"The metrics are not yet empirically validated." True. Section 4.5 labels target bands as illustrative and requiring empirical calibration. Independent validation is invited.

"The case studies are self-serving." Disclosed in Section 6 introduction and Section 7.1.

"The name AXI collides with other marks." Acknowledged in Section 1.3. We make no claim of exclusivity over the acronym.

"A linear composite score would mask catastrophic failures." Correct. AXI uses a multiplicative gated model (Section 4.5.8) precisely for this reason.

7.3 Ethics and responsible use

The Deception Guardrails (Section 4.7) and the hard CF gate (Section 4.5.8) encode ethical commitments at the framework's mathematical core rather than as separate guidelines: experience quality is not engagement maximization, consent fidelity is a hard gate, humility over overreach is non-negotiable. AXI is additive to existing safety practice, not substitutive. No AXI score exonerates a system from independent safety, fairness, or regulatory scrutiny under regimes such as the EU AI Act, NIST AI RMF, or ISO/IEC 42001 [56][57]. The Trust Ledger's append-only persistence is qualified by applicable data-rights regimes (GDPR Article 17, CCPA): user deletion requests result in cryptographic redaction of personal identifiers, not removal of the underlying audit-trail metadata required for accountability.

7.4 Open problems

The framework leaves several open research and engineering problems.

1. **Standardized benchmarks.** Cross-system benchmarks for the seven AXI metrics on a shared task battery.
2. **Cross-cultural target-band calibration.** Adjusting illustrative target bands for cultural variance without losing comparability.
3. **Telemetry proxy validation.** Empirical validation of Hesitation Time, Correction Rate, and Interruption Frequency as proxies for cognitive load.
4. **Latency budget validation.** Empirical confirmation that the 50-ms Bridge intercept budget is achievable without degrading AXI Score outcomes.
5. **Deception Guardrails under convergence.** Refinement as avatars become visually indistinguishable from humans.
6. **Continuity without surveillance.** Architectures delivering Presence Continuity without raising the data risk surface.

7. **Deferred-mode safety reasoning.** Formal safety cases for AXI in asynchronous agentic operation.
8. **Repair at population scale.** Engineering repair flows that work across tens of millions of sessions.
9. **Adversarial robustness of the gated score.** Whether the multiplicative gates can be gamed (e.g., by minimizing CF-relevant events).
10. **Adversarial users.** AXI as currently specified protects users from AI, not AI from malicious users. A user who fabricates "comfort failures" to trigger expensive repair flows, or who games their own telemetry to extract free resources, is not addressed by the framework. Bi-directional integrity is an open problem.

---

8. Conclusion

The capability of artificial intelligence has advanced rapidly. The lived experience of AI has advanced more slowly. The data summarized in Section 1.1 makes the gap visible: 46% global trust against 66% use, 95% pilot no-impact, 40%-plus forecast project cancellations.

AXI is offered as a name, a structure, and a runtime engineering specification for closing that gap. It is not a philosophy. It is a control plane, a latency budget, an API contract, a ledger schema, a gated scoring model, and a set of Deception Guardrails. We have presented two production case studies operationalizing the framework with measured metric observations across 2,650 total sessions, achieving AXI Scores of 0.91 and 0.89 — demonstrating implementability and operational measurability while explicitly disclaiming that the case studies validate the framework. We do not claim AXI is correct. We claim it is implementable, falsifiable, and improvable. We invite the research and practitioner communities to do all three.

Declarations

Competing interests. The author is the Founder and Chief Executive Officer of Sripto Corporation Private Limited, which develops both *Let Me Teach* and *SRIPTO Avatar Runtime (SAR)* — the systems discussed in Section 6. This is a material competing interest. The case studies in Section 6 are therefore presented as preliminary, non-independent operationalizations rather than as independent validation of the framework. Independent third-party replications are explicitly invited.

Funding. No external funding was received for this work.

References

[1] TechCrunch. *ChatGPT reaches 900M weekly active users.* February 27, 2026.

[2] *Sources for combined revenue and valuation figures (Section 1.1).* (a) OpenAI annualized revenue: Reuters / The Information, "OpenAI tops $25 billion in annualized revenue," March 5, 2026. https://www.reuters.com/technology/openai-tops-25-billion-annualized-revenue-the-information-reports-2026-03-05/ (b) OpenAI valuation: secondary-share transaction reporting, February 2026 (~$730B private round, with $852B target on later financing). (c) Anthropic annualized revenue (~$14B Feb 2026, ~$19B early March 2026): RootData / Sacra estimates citing The Information, March 2026. https://sacra.com/c/anthropic/ (d) Anthropic Series G valuation: Anthropic Series G announcement, February 12, 2026 (~$380B post-money). https://www.anthropic.com/news/anthropic-raises-series-g

[3] HUMAN Security, Inc. *2026 State of AI Traffic & Cyberthreat Benchmark Report.* March 26, 2026.

[4] Gillespie, N., Lockey, S., Ward, T., Macdade, A., and Hassed, G. *Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025.* KPMG and University of Melbourne, 2025 (n = 48,340 across 47 countries).

[5] Challapally, A. et al. *The GenAI Divide: State of AI in Business 2025.* MIT Project NANDA, 2025.

[6] Gartner. *Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027.* Press release, June 25, 2025.

[7] McKinsey & Company. *The State of AI: How Organizations Are Rewiring to Capture Value.* 2025 (n = 1,933 executives, 105 countries).

[8] Pew Research Center. *How the US Public and AI Experts View Artificial Intelligence.* April 3, 2025 (public n = 5,410; expert n = 1,013).

[9] Gartner. *64% of Customers Would Prefer That Companies Didn't Use AI for Customer Service.* Press release, July 9, 2024 (n = 5,728).

[10] Stanford HAI. *AI Index Report 2025.* Stanford University, 2025.

[11] S&P Global Market Intelligence / 451 Research. *Voice of the Enterprise: AI & Machine Learning, Use Cases 2025.* 2025.

[12] NIST Center for AI Standards and Innovation (CAISI). *AI Agent Standards Initiative.* National Institute of Standards and Technology, February 17, 2026.

[13] Infocomm Media Development Authority (IMDA), Singapore. *Model AI Governance Framework for Agentic AI, Version 1.0.* Launched January 22, 2026 at the World Economic Forum, Davos.

[14] Turing, A. M. Computing machinery and intelligence. *Mind*, 59(236):433-460, 1950.

[15] Legg, S. and Hutter, M. Universal intelligence: A definition of machine intelligence. *Minds and Machines*, 17(4):391-444, 2007.

[16] Bubeck, S. et al. Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:2303.12712, 2023.

[17] Yao, S. et al. ReAct: Synergizing reasoning and acting in language models. *ICLR*, 2023.

[18] Schick, T. et al. Toolformer: Language models can teach themselves to use tools. *NeurIPS*, 36, 2023.

[19] Park, J. S. et al. Generative agents: Interactive simulacra of human behavior. *UIST*, 2023.

[20] Schluntz, E. and Zhang, B. *Building Effective Agents.* Anthropic Engineering, December 19, 2024.

[21] Anthropic. *Introducing Computer Use.* October 22, 2024.

[22] OpenAI. *Introducing Operator.* January 23, 2025.

[23] Gunning, D. *Explainable Artificial Intelligence (XAI).* DARPA Technical Report, 2017.

[24] Arrieta, A. B. et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82-115, 2020.

[25] Ribeiro, M. T., Singh, S., and Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. *ACM SIGKDD*, 22:1135-1144, 2016.

[26] Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. *NeurIPS*, 30:4765-4774, 2017.

[27] Passi, S. and Vorvoreanu, M. *Overreliance on AI: Literature Review.* Microsoft Research / Microsoft Aether, June 2022.

[28] Sperrle, F. et al. Tell me something that will help me trust you: A survey of trust calibration in human-agent interaction. arXiv:2205.02987, 2022.

[29] Henrique, B. M. and Santos, E. Jr. Dynamic trust calibration using contextual bandits. arXiv:2509.23497, 2025.

[30] Radford, A. et al. Learning transferable visual models from natural language supervision. *ICML*, 38, 2021.

[31] Alayrac, J.-B. et al. Flamingo: A visual language model for few-shot learning. *NeurIPS*, 35, 2022.

[32] Liu, H., Li, C., Wu, Q., and Lee, Y. J. Visual instruction tuning. *NeurIPS*, 36, 2023.

[33] OpenAI. *GPT-4V(ision) System Card.* 2023.

[34] Gemini Team, Google. *Gemini: A Family of Highly Capable Multimodal Models.* arXiv:2312.11805, 2023.

[35] Brooks, R. A. Intelligence without representation. *Artificial Intelligence*, 47(1-3):139-159, 1991.

[36] Liu, Y. et al. *Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI.* arXiv:2407.06886, 2024.

[37] New Market Pitch. *Humanoid Robotics Market Funding Trends (2022–2026).* Industry-data report. https://newmarketpitch.com/blogs/news/humanoid-robotics-funding-trends (freely accessible; aggregates 57 equity deals 2022–2025).

[38] Norman, D. A. *The Design of Everyday Things*, revised and expanded edition. Basic Books, 2013.

[39] Norman, D. A. *Emotional Design.* Basic Books, 2004.

[40] Shneiderman, B. *Designing the User Interface.* Addison-Wesley, 1987.

[41] Shneiderman, B. *Human-Centered AI.* Oxford University Press, 2022.

[42] Nielsen, J. Enhancing the explanatory power of usability heuristics. *CHI*, pages 152-158, 1994.

[43] Lombard, M. and Ditton, T. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[44] Slater, M. and Wilbur, S. A framework for immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603-616, 1997.

[45] Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience.* Harper & Row, 1990.

[46] Lee, J. D. and See, K. A. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50-80, 2004.

[47] Nielsen Norman Group. *AI-Era Interaction Patterns.* 2024.

[48] Weizenbaum, J. ELIZA — a computer program for the study of natural language communication. *Communications of the ACM*, 9(1):36-45, 1966.

[49] Picard, R. W. *Affective Computing.* MIT Press, 1997.

[50] Breazeal, C. *Designing Sociable Robots.* MIT Press, 2002.

[51] Ayers, J. W. et al. Comparing physician and AI chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, 183(6):589-596, 2023. (The study measured *perceived empathy by third-party text evaluators*, not real-time empathetic interaction.)

[52] Amershi, S. et al. Guidelines for human-AI interaction. *CHI*, paper 3, 1-13, 2019.

[53] Google PAIR. *People + AI Guidebook.* 2019 (continuously updated).

[54] Xu, W. *A "User Experience 3.0 (UX 3.0)" Paradigm Framework.* arXiv:2403.01609, 2024.

[55] Industry writing on "Agent Experience (AX)" and "Artificial Experience (AX)," 2025-2026.

[56] EU AI Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council. Official Journal of the European Union, 2024.

[57] ISO/IEC 42001. *Information Technology — Artificial Intelligence — Management System.* International Organization for Standardization, 2023.

[58] Yeluri S. D. S. Sri Vardhan. Effects of business intelligence tools on financial performance of IT industry. *AIP Conference Proceedings*, 2971, 020036, 2024. https://doi.org/10.1063/5.0196170

[59] Vectara. *Hallucination Leaderboard (HHEM Benchmark).* 2026.

[60] Suleyman, M. *Published commentary on "Seemingly Conscious AI."* 2025.

[61] Bailenson, J. N. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior*, 2(1), 2021.

[62] Mitra, A. and Sethuraman, S. C. *The Qey: Implementation and Performance Study of Post-Quantum Cryptography in FIDO2.* arXiv:2510.21353, 2025.

[63] Mitra, A. and Sethuraman, S. C. *Verifiable Passkey: The Decentralized Authentication Standard.* arXiv:2512.21663, 2025.

[64] Nielsen, J. *Usability Engineering.* Morgan Kaufmann, 1993 (response-time thresholds: 0.1 s, 1 s, 10 s).

[65] Reeves, B. and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, 1996.

[66] Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021*, 610–623.

Appendix A: Glossary of AXI Terms

AXI (Artificial eXperience Intelligence). The engineering discipline defined in this paper.

AXI Bridge. The runtime control plane mediating between AI capability and human experience. Layer 5 of the AXI Stack.

AXI Stack. The five-layer reference architecture: Perception, Cognition, Action, Perceptual Substrate, Experience (AXI Bridge).

AXI Principles. Nine design priorities for the experience layer (Section 4.4).

AXI Metric Suite. Seven metrics with explicit telemetry-based definitions (Section 4.5).

AXI Score. The multiplicative gated composite (Section 4.5.8).

Perceptual Substrate. Layer 4. The substrate through which the AI is perceived — voice, avatar, kiosk, hologram, robot. Deliberately broader than Brooks-strict embodiment.

Pacing Controller, Transparency Orchestrator, Comfort Monitor, Trust Ledger, Consent Broker, Continuity Manager, Repair Orchestrator. The seven Bridge components (Section 4.2.5).

Commitment-class facts. User-stated preferences, system-issued promises, user-issued corrections, user-confirmed boundaries — recorded in the Trust Ledger, distinct from RAG retrieval-class facts.

ANBE (Asynchronous Non-Blocking Evaluation). Bridge operating mode for sub-millisecond inference engines (Section 4.3.2).

Mandate envelope. The pre-authorized scope, boundaries, and notification triggers for deferred-mode operation (Section 4.6).

Mandate Rollback Protocol (MRP). Termination protocol for expired/revoked mandates mid-transaction (Section 4.6).

Deception Guardrails. Seven hard requirements preventing anthropomorphic dark patterns (Section 4.7).

TtC, II, PC, TDR, RE, CLI, CF. The seven AXI Metrics defined in Section 4.5.

Appendix B: Measurement Methods

This appendix documents the measurement instruments used to derive the values reported in Tables 6 and 7. Independent replications should report the equivalent instruments to support direct comparison.

Session definition. A *session* is defined as a continuous user-system interaction bounded by either (a) explicit user termination (log-out, close, "end session" intent), or (b) 15 minutes of bidirectional inactivity, whichever comes first. Multi-turn interactions interrupted by gaps of less than 15 minutes are treated as one session. This definition is held constant across both case studies.

Time-to-Comfort (TtC) — measurement instrument. The "first user-task statement" is detected by an intent classifier (fine-tuned encoder on a labeled corpus of in-domain task-intent utterances). Task-relevant chunks are tagged by the same classifier. User Hesitation Time is the wall-clock duration between successive user turns, measured at the input boundary. Correction Rate is the count of user corrections per minute, where a correction is detected by a separate classifier trained on lexical (e.g., "no," "actually," "I meant") and structural (e.g., re-issued tool call with edited arguments) patterns. The 10-minute sliding window aggregates these signals.

Interaction Integrity (II) — judge configuration. Adversarial probes are scored by an LLM-as-judge ensemble: two frontier judges (Claude and GPT-class) score each probe independently, and disagreement above a configured threshold triggers a manual-audit fallback on 1% of disagreements. The probe set is rotated daily from a pool of approximately 5,000 adversarial probes spanning Accuracy, Calibration, and Scope. Minimum 50 probes per session. Static probe sets are forbidden.

Presence Continuity (PC) — audit protocol. Session-pair audits inject probe questions into sessions 2 through 5 that test whether commitment-class facts asserted in session 1 are honored. PC measures whether the system *acted on* commitments, not merely whether it retrieved them. The audit covers preferences, promises, corrections, and confirmed boundaries.

Trust Decay Rate (TDR) — incident tagging. Critical incidents are tagged by the Repair Orchestrator using objective rules (user-issued correction; user-tagged frustration via affect classifier; system-confessed error). Microsurvey is gated by both (a) at least 7 days since the last trust survey to this user AND (b) at least 50 interaction turns since the last trust survey.

Repair Efficacy (RE) — failure classifier. Detected experience failures are identified by a learned classifier whose precision and recall on a held-out validation set must be reported alongside RE for AXI conformance. Recovery is measured by passive comfort telemetry (Hesitation Time return to baseline, Correction Rate decrease) within three turns post-repair.

Cognitive Load Index (CLI) — telemetry. User Hesitation Time, Correction Rate, and Interruption Frequency are measured passively at the I/O boundary. Response Density is system-emitted words per second weighted by syntactic complexity (computed via dependency-tree depth on the emitted text). NASA-TLX is deployed once per session maximum.

Consent Fidelity (CF) — audit. The Consent Broker ledger is fully automated. Action Layer events are audited 100%. Perception Layer events are sampled (the 2% tolerance accommodates measurable noise floor; see Section 4.5.7). Identity-claim events are audited 100%.

Cohort composition. *Let Me Teach* (n = 2,500) sessions are drawn from active users of the launched product (https://letmeteach.in) during March–April 2026, excluding internal Sripto accounts. *SAR* (n = 150) sessions are drawn from a controlled pre-launch pilot deployment with consented external participants in receptionist-mode; SAR is not yet publicly launched.

Variance estimates. Variance estimates (IQR, SD, 95% CI) are omitted in this preliminary report. Independent replications, which the framework explicitly invites, should include confidence intervals appropriate to their sample design.

Reproducibility scope. The measurement instruments above describe what *was used* in the production telemetry of the case-study systems. They are documented here to support reproducibility, not to standardize across implementations. Independent reference implementations may use different but equivalent instruments, provided the resulting AXI metric values are comparable in interpretation.