

The Water We Swim In

A Candidate Environmental Partition for Persistent Human–AI Interaction

Perspective and Pre-Pilot Research Agenda

Micheal Charles Preble

ORCID 0009-0005-3906-6028

Operant Dyad Research Program

SEPTEMBER 13, 2026

PUBLICATION NOTE

Independent research publication. Not affiliated with or endorsed by Stanford University or the Stanford Institute for Human-Centered Artificial Intelligence. Visual presentation is inspired by public academic white-paper conventions used by Stanford HAI and similar research institutes.

Publication Status

Manuscript status	Perspective / conceptual-methodological proposal
Empirical status	Candidate partition not validated
Technical companion	OD-ENV-01 - Ambient Condition Coding and Manipulation
Technical status	PRE-PILOT / HELD AT PROVENANCE GATE
Revision	v0.3 - Pre-publication revision; internal adversarial review complete
License	CC BY 4.0

PROTECTED LINE

Ecotone can fire and Water can still lose.

Epistemic Status Key

Marker	Meaning
[EXISTING EVIDENCE]	Claim supported by cited empirical or established scholarly work
[PROPOSED PARTITION]	Candidate analytical distinction introduced by this paper
[ILLUSTRATIVE SCENARIO]	Example used to clarify a proposed distinction; not presented as observed evidence
[PLANNED TEST]	Procedure specified for future empirical evaluation
[UNVERIFIED ASSUMPTION]	Possibility retained for testing but not treated as established

These markers prevent conceptual propositions, observed findings, illustrative cases, and planned tests from being read as having the same evidentiary status. Established frameworks are not background citations to be acknowledged and moved past; they are competitors that are allowed to make the proposed technical vocabulary unnecessary.

Abstract

Persistent human-AI interaction increasingly involves more than a prompt and a response. Memory systems, retrieval, personalization, permissions, routing, interface defaults, organizational policy, professional obligations, and accumulated interaction history can shape later behavior in some deployed systems. Existing traditions - including sociotechnical systems, human factors, distributed cognition, human-AI teaming, personalization, user modeling, automation bias, and AI governance - already explain substantial portions of this territory, and this paper does not claim that environmental influence itself is novel.

[PROPOSED PARTITION] This paper introduces The Water as an attentional metaphor and proposes a candidate provenance-and-control-surface partition between inherited human conditions and engineered synthetic conditions in persistent human-AI interaction. The proposed value of the partition is not that "social" and "technical" factors have never been separated, but that persistent generative systems can carry personalized, stateful, partly opaque conditions across interaction episodes while humans simultaneously remain embedded in institutional, professional, economic, and social conditions. The paper further restricts Ecotone to observable activated boundary processes in which documented human and synthetic conditions become jointly

consequential through invocation, contestation, translation, gating, propagation, override, refusal, reconciliation, or transformation.

[PLANNED TEST] Three domains motivate the research agenda: memory and context sovereignty, practical authority and meaningful oversight, and the difference between immediate assisted performance and durable human capability. The architecture has an explicit falsifier: if strong existing frameworks classify difficult real episodes as reliably as the candidate partition and preserve the same causal, governance, authority, intervention, measurement, or predictive distinctions, the proposed technical vocabulary has not earned separate status. OD-ENV-01 is designed to test that possibility adversarially, but no validation result is claimed here; its first two-thread shakedown remains held until thread-scoped contemporaneous provenance can be established without inference.

Keywords: persistent human-AI interaction; sociotechnical systems; AI memory; human agency; human oversight; personalization; context governance; Relationship State; cognitive offloading; human-AI collaboration

1. From Prompt to Configured Interaction

[EXISTING EVIDENCE] Early public interaction with large language models was often reasonably approximated as a sequence of bounded exchanges: a user supplied a prompt, a model generated an output, and evaluation focused on properties such as accuracy, usefulness, safety, bias, trust, or reliance. That representation becomes less complete when deployed systems preserve user information across sessions, retrieve project material, route queries through multiple components, act through tools, apply persistent preferences, or operate within institutional workflows. Stanford's Digital Economy Lab now describes persistent personal context as an emerging infrastructure problem and argues for user-governed context that can persist across interactions, move across tools, and be shared selectively by purpose and consent.

The change does not mean that every contemporary AI interaction has become persistent or relational. Many remain episodic, stateless, or weakly stateful, and the present paper should not be read as a prevalence claim about all AI use. The narrower research-position claim is that some questions about persistent systems can make the surrounding configuration a necessary unit of analysis, because claim-relevant conditions may reside in memory state, routing, permissions, system configuration, institutional policy, professional obligation, or prior interaction. The Operant Dyad framework addresses one part of this shift by treating repeated human-AI work as potentially adaptive and by separating accumulated Relationship State from individual prompts; this manuscript instead asks about the conditions surrounding that relationship.

2. Initial Ecological Scope

This manuscript uses the broad phrase persistent human-AI interaction, but the first empirical target is intentionally narrower. OD-ENV-01 is initially concerned with persistent, work-like, research, project, and capability-bearing interactions in which memory, authority, correction, oversight, or continuing context can materially affect subsequent activity. This includes independent professional work as well as institutionally constrained work, provided that the relevant environmental state can be documented rather than inferred. The scope is therefore narrower than the title's full conceptual horizon.

A clinician, student, software developer, independent researcher, caseworker, customer-support worker, and AI-companion user do not occupy interchangeable ecologies. They differ in liability, labor power, data ownership, institutional oversight, access to system configuration, error costs, and ability to refuse or exit. The same confirmation gate can therefore have radically different meaning depending on who bears the cost of delay, who is accountable for error, and who controls the surrounding system. The first research corpus should preserve those differences instead of treating "the user" as a socially unlocated participant.

Clinical care, education involving minors, intimate AI companionship, and other especially sensitive domains may inform the literature synthesis but are not automatically eligible for the initial real-thread corpus. Separate consent, privacy, and domain-specific review requirements would be necessary before treating those interactions as empirical study material. This prevents a broad conceptual paper from silently assuming that one future codebook is ethically or analytically portable across every form of human-AI life. It also preserves the distinction

between an agenda's conceptual scope and the actual populations a first empirical program can responsibly study.

3. Why "Context" Is Not the Claim

[EXISTING EVIDENCE] Sociotechnical scholarship has long rejected the idea that technologies can be understood independently of institutions, roles, work practices, and social organization. Human factors has documented automation misuse, complacency, supervisory-control problems, workload, reliance, and skill effects for decades (Bainbridge, 1983; Parasuraman & Riley, 1997; Lee & See, 2004). Distributed and extended cognition have likewise treated tools and external representations as components of cognitive activity rather than neutral objects outside it (Hutchins, 1995; Clark & Chalmers, 1998; Risko & Gilbert, 2016). Any claim that human-AI research simply forgot environmental context would therefore be historically and analytically indefensible.

Contemporary human-AI teaming research makes those dependencies explicit rather than treating model capability as sufficient. Gonzalez and colleagues' 2026 PNAS Nexus perspective integrates reasoning, memory, attention, trust calibration, shared mental models, task structure, training, and organizational factors into a complementarity framework for human-AI teams. Zhang, Wang, and Yi's scoping review of 134 top-tier HCI/CSCW papers similarly documents a substantial existing research vocabulary around agency configurations, control mechanisms, and interaction contexts. The Water argument therefore cannot be that surrounding conditions constitute a newly discovered category of HAI.

The research question is narrower: does persistent generative AI make a provenance-and-control-surface partition sufficiently useful to improve empirical discrimination, accountability assignment, intervention design, or longitudinal measurement in some difficult cases? If the answer is no, sociotechnical systems, automation bias, personalization, agency theory, human factors, memory governance, and related established concepts remain sufficient. That possibility is built into the empirical design rather than added as a rhetorical caveat. The paper's contribution therefore lies in proposing a contestable cut and specifying how that cut is allowed to disappear.

4. The Water as Attentional Metaphor

[PROPOSED PARTITION] The Water is a working metaphor for conditions that become difficult to notice because they form the ordinary environment of persistent human-AI interaction. It is not a latent variable, universal mechanism, or claim that memory policies, labor incentives, interface defaults, organizational hierarchy, professional norms, personalization, and cultural assumptions are one causal object. Its job is to direct attention toward conditions that may otherwise remain backgrounded long enough to determine which more precise objects are actually present. If that attentional function turns into an all-purpose explanation, the metaphor has failed.

Normalized assumptions can structure action without appearing as explicit decisions. An institution can reward throughput for years before an AI confirmation gate makes consequences of that metric newly visible. A user can assume that separate projects remain compartmentalized until a persistent personalization system applies material across those boundaries. A professional can formally retain final authority while workflow speed, information asymmetry, or verification cost makes exercising that authority increasingly difficult.

These examples are illustrative scenarios, not findings of this paper. Their purpose is to identify the kind of source material the empirical program would eventually have to discriminate. A good Water example is therefore not one that makes the metaphor sound compelling; it is one that can be decomposed into observable conditions and then challenged by stronger existing explanations. The technical work should move from metaphor toward separable evidence, not from metaphor toward increasingly comprehensive theory.

5. Candidate Provenance-and-Control-Surface Partition

[PROPOSED PARTITION] The paper proposes two top-level provenance classes: ENV-HUMAN and ENV-SYNTH. They should not be read as proven natural kinds or as mutually exclusive labels for whole episodes.

The proposed partition instead asks where a consequential condition entered the episode and which control surface must be altered if that condition is to change. Hybrid cases are expected rather than treated as anomalies that must be forced into one bin.

ENV-HUMAN refers provisionally to active institutional, professional, social, historical, economic, cultural, or role-based conditions brought into an interaction from the human setting. Examples include throughput metrics, managerial hierarchy, professional documentation rules, deadlines, accountability requirements, liability structures, economic incentives, or prior human commitments. A workplace existing in the background is insufficient; the condition must constrain, demand, or license a move observable in the episode. Participant-stated requirements may constitute evidence, but they are not automatically equivalent to independently documented institutional policy.

ENV-SYNTH refers provisionally to active engineered system conditions such as memory configuration, system instruction, routing, retrieval, permission scope, provenance display, interface default, persistence behavior, evaluation target, tool access, model policy, or termination rule. A model merely possessing memory or tool capability is insufficient; the condition must constrain, demand, or license a move observable in the episode. A provider update with no locally visible configuration may still be synthetic in origin when it is independently documented, but undocumented behavioral drift should remain unavailable for confident coding. The top-level category is therefore a provenance class, not a causal mechanism.

Question	ENV-HUMAN	ENV-SYNTH
Where did the condition primarily enter the episode?	Institution, role, profession, market, law, culture, prior human arrangement	Product, model, memory layer, interface, routing, permission, retrieval, system policy
What evidence would establish it?	Policy, role record, contract, workflow rule, contemporaneous participant statement	Configuration record, system state, UI capture, tool scope, model/version, retrieval or permission trace
Who can usually intervene?	Employer, institution, regulator, profession, team, individual human actor	Provider, deployer, administrator, product designer, system owner, sometimes user
What does the label not explain?	The technical mechanism enforcing or translating the condition	The human purpose, legitimacy, labor relation, or institutional consequence behind it
Mixed-origin case	Human rule may originate institutionally and later be technically enforced	Synthetic implementation may encode a human-authored institutional demand

The partition earns technical value only if separating these provenance classes changes an actual analysis. A difference in terminology is insufficient. A difference in intervention owner, evidence demand, authority assignment, accountability pathway, manipulability, measurement, or prediction would be more consequential. If a conventional sociotechnical account already makes those same distinctions at the same grain, the candidate partition is redundant.

6. Synthetic Conditions Without Recreating "Context"

ENV-SYNTH can become another undifferentiated bucket if model behavior, memory, routing, permissions, interface design, organizational configuration, and provider policy are treated as interchangeable. That would reproduce at the synthetic level the same problem the manuscript attributes to overly broad uses of context. The first-level label should therefore be interpreted only as a provenance class. Causal explanation must remain at the most specific documented level the evidence permits.

For future full-run analysis, synthetic conditions are expected to be described as memory, retrieval, system instruction, routing, permission, interface default, tool access, evaluation target, safety/termination rule, provenance display, or another specifically documented mechanism. These are reporting dimensions inside the existing protocol rather than a new theoretical layer. The frozen two-thread packet does not receive a seventh instrument or a new construct because a reviewer correctly noticed heterogeneity inside ENV-SYNTH. Instead, the record must show what specific synthetic condition is actually evidenced.

This distinction prevents several causal errors. If an assistant behaves differently after a provider update, the relevant explanation might be model/version drift rather than Relationship State. If a stored preference exists but was not demonstrably retrieved, storage alone should not be used to explain behavior. If hidden routing changes the responding model, a coder should not attribute the result to the human-AI relationship simply because the visible interface appears continuous. Unknown system conditions therefore reduce the available code set instead of being completed by inference.

7. Ecotone: Activated Boundary Process, Not Third State

[PROPOSED PARTITION] Ecotone is restricted to an activated boundary process. It is not the interface in general, not a third state between human and machine, and not a synonym for sociotechnical interaction. The proposed label applies only when documented human and synthetic conditions become jointly consequential through an observable activation. This restriction is what prevents the term from absorbing every episode in which a person happens to use an AI system.

For the planned empirical program, an Ecotone candidate requires four elements: H, an active source-supported human condition; S, an active source-supported synthetic condition; A, an observable activation involving invocation, contestation, translation, gating, propagation, override, refusal, reconciliation, or transformation; and C, a documented consequence for action, available options, authority, scope, system state, or decision path. In compact form: $ECO = H \wedge S \wedge A \wedge C$. Mere coexistence does not qualify, and mere verbal reference to policy or memory without a consequential change does not qualify either.

A memory bug affecting a user without any evidenced human-condition crossing may be an ordinary product defect rather than Ecotone. A workplace interaction in which organizational policy never becomes consequential is not Ecotone merely because an institution exists. A refusal produced entirely by a system policy with no observable human/synthetic boundary event may remain a synthetic-system episode or be completely described by an existing framework. The proposed unit is therefore event-like rather than atmospheric.

7.1 Constructed Coding-Grain Specimen

[ILLUSTRATIVE SCENARIO - CONSTRUCTED, NOT OD-ENV-01 DATA] The following specimen exists only to show the grain at which the proposed code would have to operate. It was constructed from the confirmation-button × throughput scenario already used in this manuscript and is not an observed transcript, empirical result, or eligible OD-ENV-01 record.

thread_context	Enterprise claims-review workflow
human_condition	Same-day closure metric documented by employer
synthetic_condition	AI-generated recommendation with one-click APPROVE AND SUBMIT; source panel available but closed by default
activation_anchor	System presents submission recommendation while same-day closure target remains binding
observable_act	Worker clicks APPROVE AND SUBMIT after approximately three seconds without opening provenance panel
consequence	Claim moves from review state to submitted state
ENV-HUMAN	Fires: documented closure metric constrains the practical review window
ENV-SYNTH	Fires: interface default and confirmation path structure the available action
ECO	Fires provisionally because $H + S + A + C$ are all present
DYAD	Does not fire: materially the same event could occur in a first-session stateless interaction
RS	Does not fire: no accumulated relational configuration is required
adversary_first_code	Automation bias + choice architecture + sociotechnical workflow analysis

thread_context	Enterprise claims-review workflow
closest_existing_explains	Fast acceptance of automated recommendation under workflow pressure and interface design
Water_residual_candidate	Separation of intervention owners: employer throughput metric versus deployer/interface confirmation design
residual_test	Does ordinary STS/human-factors analysis already preserve those two intervention owners?
terminal_judgment_for_specimen	EXISTING-SUFFICES
reason	A competent sociotechnical/human-factors analysis can already identify both organizational pressure and interface design and assign interventions to each without losing a governance-relevant distinction

The specimen is useful precisely because the Water coding can fire while the terminal judgment remains EXISTING-SUFFICES. Ecotone therefore does not function as a victory condition. A boundary event may be recognizable under the proposed vocabulary while providing no residual analytical advantage over existing theory. That is the behavior the later adversarial protocol must be capable of preserving.

8. Relationship to Operant Dyad, Relationship State, Memory, and Personalization

The Operant Dyad remains the persistent adaptive human-AI working relationship. Relationship State remains accumulated relational configuration that materially changes subsequent interaction. The present paper concerns conditions surrounding those objects and should not collapse environment, memory mechanism, relationship, and accumulated state into a single architecture. Each occupies a different analytical level.

A stored preference by itself is not Relationship State. A retrieval event using that preference is primarily evidence of a synthetic memory process. A long transcript is not Relationship State merely because prior text is available, and cross-session continuity is not sufficient without evidence that an accumulated relational configuration materially changes what happens. Ordinary persistence should therefore not be promoted into relational state merely because the system remembers.

A confirmed correction that later changes interpretation, scope, authority, or task behavior may provide stronger evidence of Relationship State, particularly when its continuing force depends on the prior working relationship rather than generic personalization. Conversely, a provider-generated user profile that changes answers without a relationship-specific confirmation history may be better treated as personalization or user modeling. When provenance cannot decide among these accounts, AMBIG is the correct result. The codebook must not complete a relationship from plausibility.

The working decision procedure is therefore conservative. Stored or retrieved information begins as a synthetic-system candidate, not as Relationship State. Prior interaction appearing again establishes continuity but not necessarily relational accumulation. RS becomes a candidate only when an earlier relational configuration is traceably available and materially changes the current episode. If ordinary personalization fully explains the same event, EXISTING-SUFFICES remains available.

9. Domain I: Memory and Context Sovereignty

Stanford's 2026 portable-memory work frames persistent personal context as both a capability and a governance problem. Its stated goal is a user-governed layer that can persist across interactions, move across tools, and be shared selectively by role, consent, and purpose rather than becoming trapped inside proprietary cognitive layers. This work supports the relevance of portability, persistence, scope, and user control. It does not validate the Water partition.

Persistent memory raises questions that cannot be reduced to storage accuracy alone. Scope, provenance, relevance, correction, revocation, compartmentalization, portability, and authority matter because a remembered item can be true yet inappropriate for a current task. Users may reasonably want one persistent project history while expecting another domain to remain unaffected. A system that retains more information is therefore not automatically a system that governs context better.

Fang et al.'s 2026 DRIFTLens paper directly supports a bounded version of this concern. Across four LLMs and ten user-attribute categories, the researchers report medium-to-large memory-induced changes in expressed reasoning trajectories relative to no-memory baselines even when final responses remain fluent, on-topic, and plausible. The result does not establish harmful personalization generally, nor does it show that every stored attribute influences every task. It demonstrates that injected user memory can measurably alter reasoning rather than merely surface wording.

The Water perspective asks a governance question rather than claiming a new memory mechanism: which effects belong to accumulated relationship state, and which belong to the engineered environment carrying, retrieving, or applying that state? The answer could affect portability, correction, responsibility, and intervention. Yet if existing memory-governance and personalization frameworks already represent those distinctions without consequential loss, Water adds nothing technical. That comparison remains open.

10. Domain II: Practical Authority and Meaningful Oversight

Human-in-the-loop governance often assumes that placing a person at an approval point preserves meaningful control. Human-factors research has long shown that formal supervisory placement is not equivalent to effective supervision, especially under workload, overreliance, information asymmetry, habituation, or poor interface design (Bainbridge, 1983; Parasuraman & Riley, 1997; Lee & See, 2004). The proposed Water analysis does not replace those findings. Its narrower concern is whether separating provenance and control surfaces improves accountability or intervention in persistent-AI settings.

A worker may technically be able to reject an AI recommendation while facing throughput metrics, managerial expectations, difficult provenance access, verification fatigue, or liability structures that make meaningful challenge expensive. A confirmation mechanism can therefore be formally present while the surrounding human and synthetic conditions jointly make its use largely ceremonial. The constructed specimen in §7.1 shows how such an episode might be coded while still being conceded to existing human-factors and STS analysis. That concession is a feature of the method rather than an embarrassment.

Human-AI teaming literature already treats role partition, trust calibration, task structure, knowledge infrastructure, attention, and accountability as core determinants of complementarity. The candidate partition must therefore demonstrate more than the proposition that institutions and interfaces interact. Its burden is to show whether provenance-specific separation makes authority conflicts more traceable, experimentally manipulable, or governable than established approaches already do. A failure to clear that burden should produce EXISTING-SUFFICES.

11. Domain III: Assisted Performance and Durable Human Capability

Immediate task performance and durable human capability should not be assumed to be identical outcomes. AI systems can improve assisted performance while affecting persistence, independent skill, verification ability, self-efficacy, or professional judgment in different directions. Those effects must be measured separately rather than aggregated into a claim that AI either strengthens or weakens people. This paper therefore uses empirical studies only for the specific outcomes they measured.

Liu, Christian, Dumbalska, Bakker, and Dubey report randomized human-AI experiments involving 1,222 participants across tasks including mathematical reasoning and reading comprehension. Their study found improved assisted performance alongside lower subsequent unassisted performance and reduced persistence, with effects emerging after approximately ten minutes of exposure in the tested settings. Public study materials further indicate that these effects were particularly associated with interactions in which users requested direct

solutions rather than hints, sharpening the distinction between substitution and scaffolding. The finding does not establish generalized human deskilling.

Ehsan and colleagues report a year-long radiation-oncology deployment involving 42 participants, including 24 interviews, five workshops, and 52 think-aloud sessions. Their work describes early operational gains alongside what the authors call "intuition rust," or gradual dulling of expert judgment, and uses those observations to motivate dignified human-AI interaction. This is a different evidentiary grain from Liu et al.'s randomized short-exposure study. The paper therefore does not combine persistence, unassisted performance, and expert intuition into one measured construct.

The measurement distinction matters. Immediate assisted performance concerns output under support; unassisted retention concerns later performance without support; persistence concerns willingness to continue difficult work; verification skill concerns detecting weaknesses in AI output; professional judgment concerns domain-specific interpretive competence; and self-efficacy concerns belief in one's own ability. The Water research question is conditional: which environmental configurations make assistance function as substitution, scaffolding, verification aid, dependency amplifier, or capability-supporting partner?

12. Prior-Art Audit: Procedure, Examples, and Limits

The present paper does not claim to contain a systematic review. A rapid adversarial scoping audit was conducted to identify the strongest nearby traditions and to ask whether the proposed Water architecture could be reduced to them. Search and review work on September 13, 2026 included academic discovery tools, direct publisher records, ACM/CHI materials, arXiv, PNAS Nexus, Stanford Digital Economy Lab materials, and targeted review of adjacent human-factors and cognition traditions. The purpose was falsification-oriented orientation rather than exhaustive coverage.

Query families included combinations of persistent human-AI interaction, human-AI teaming, sociotechnical AI, human AI agency, AI memory personalization, persistent personal context, automation bias human oversight, cognitive offloading AI, AI skill atrophy, memory reasoning drift, portable AI memory, agent permissions authority, and human-AI collaboration longitudinal. Inclusion prioritized work directly addressing persistent state, memory or personalization, authority and control, human capability, sociotechnical coordination, or relevant governance. Generic AI-ethics commentary, product marketing without research content, unrelated stateless benchmarks, and sources lacking adequate traceability were not treated as substantive evidence.

No numerical novelty score was computed. An earlier working phrase - MODERATE / DEFENSIBLE GAP - has therefore been removed from the manuscript rather than presented as though produced by a formal review instrument. The defensible conclusion is narrower: the conceptual neighborhood is crowded, and a residual empirical question remains unresolved about whether provenance-specific separation retains analytical or governance information that strong existing frameworks lose.

12.1 Three Audit Items

The following examples make the adversarial logic inspectable. They are examples from the scoping audit rather than empirical OD-ENV-01 episode classifications.

Audit item	Existing framework	What existing work already carries	Water status
Human and technical conditions shape workflow outcomes	Sociotechnical systems	Joint optimization, institutional organization, technical subsystem, work practice, intervention at multiple levels	EXISTING-SUFFICES for the broad proposition. Water has no residual merely by renaming social/technical provenance.
Confirmation under throughput pressure	Human factors + automation bias + choice architecture + STS	Reliance, supervisory control, workload, interface effects, organizational pressure, intervention opportunities	EXISTING-SUFFICES in the constructed \$7.1 specimen. Water does not retain unique governance information there.
Persistent user memory changes later reasoning	Personalization / user modeling / persistent-memory research, including DRIFTLENS	Memory-induced reasoning drift, retrieval effects, user-state conditioning, personalization risk	MECHANISM ALREADY EXPLAINED. The unresolved Water question is governance decomposition - relationship state versus memory machinery, scope, intervention owner, correction, and portability - not the existence of memory effects.

These examples intentionally do not manufacture a Water residual. The audit establishes that substantial parts of the proposed territory are already well occupied. A residual can only be earned on an episode where removing the proposed partition loses a consequence involving cause, governance, authority, intervention, measurement, or prediction. Until such an episode survives adversarial coding, the manuscript claims only a candidate research program.

The principal rival stack retained for direct comparison is intentionally small: sociotechnical systems; human factors and automation bias; human-AI teaming; personalization and persistent-memory research; and distributed or extended cognition plus agency and governance scholarship. Adding dozens of neighboring literatures would make the contest less interpretable rather than more rigorous. The future adversary is allowed to introduce another established framework only by naming it explicitly. The comparison therefore remains open without becoming theoretically unbounded.

13. The Falsifier, Operationally Bounded

The central falsifier remains: if strong existing frameworks classify difficult real episodes as reliably as the candidate Water partition and preserve the same consequential information, the Water architecture has not earned separate technical status.

"Consequential information" is restricted to six forms of difference: cause attribution, governance responsibility, authority allocation, intervention owner, measurement requirement, and prediction. A Water classification does not earn residual value because it is clearer, more memorable, more integrative, or better suited to a public essay. At least one of those six analytical consequences must change relative to the strongest established account. Otherwise the appropriate result is EXISTING-SUFFICES.

The two-thread shakedown does not test this falsifier quantitatively. Its purpose is procedural: determine whether provenance can be completed, independent segmentation produces usable episodes, both coders can use the shared form, ECO does not become atmospheric interaction, RS does not become generic continuity, and EXISTING-SUFFICES can compete cleanly. No construct-survival claim is permitted from the shakedown. A procedurally successful pilot can therefore return no Water residuals at all.

If the packet survives the shakedown, the planned 8-12-thread run will require a preregistered analysis plan before full coding begins. That plan must specify the inter-rater statistic, reliability threshold, treatment of AMBIG, minimum evidentiary standard for WATER-RESIDUAL, and program-level PASS / FAIL / INDETERMINATE rule. Those thresholds will not be selected from shakedown outcomes. The larger run should therefore begin only after procedural behavior is known but before construct-survival outcomes are available.

14. OD-ENV-01: Current Empirical Status

OD-ENV-01 is not presently a validated construct-discrimination protocol. It is a runnable pre-pilot shakedown packet whose immediate purpose is to discover whether the procedure breaks before a larger corpus is collected. The six frozen components are the prior-art adversary instruction, Water-side coder instruction, shared episode record, disagreement table plus pilot gate, independent segmenter instruction, and provenance slip. The current technical companion is versioned separately from this manuscript.

The run order is fixed. Provenance slips are locked first; an individual who is neither coder independently identifies activation-anchored spans before either theory-specific coding pass; those spans are then locked and released identically to the Water-side coder and prior-art adversary. Both work independently and do not reconcile first-pass disagreement. A thread can produce zero, one, or multiple episode spans, while coding and judgment occur at the episode level.

Eligible threads must be real persistent interactions, have full transcripts rather than reconstructed demonstrations, permit the provenance slip to be completed without guessing claim-relevant conditions, contain a non-theoretical activation anchor identifiable by the segmenter, and contain enough observable friction for EXISTING-SUFFICES to be a genuine possibility. Reconstructed examples, excerpts hiding the activation point, or threads requiring inference about hidden memory, permission, routing, or institutional conditions are ineligible. The constructed specimen in §7.1 is explicitly excluded.

The shakedown remains HELD AT THE PROVENANCE GATE. Candidate archived threads were not promoted because environment-level evidence of persistent memory did not establish the thread-scoped memory state required by the slip. This hold is part of the method rather than an administrative delay: an environmental study cannot infer the environment it proposes to classify. If eligible archival material does not exist, future ordinary work can become eligible only when relevant settings are recorded contemporaneously and the interaction was not initiated to exhibit Water.

15. Model and System Accountability

Any future empirical test must document system conditions relevant to the claim being made. At minimum, the provenance record should establish product, model or version where known, deployment owner, time range, thread-scoped memory setting, tools enabled, override structure, known institutional requirements, known interface defaults, and whether relevant routing, retrieval, permissions, system instructions, or configuration changes are documented. Unknown fields remain unknown rather than being promoted through behavioral inference. A condition cannot support a code simply because the resulting output is compatible with that condition having existed.

This reporting posture is consistent with Navarro, Syriani, and Arawjo (2026), who argue that LLM-integrated HCI studies should report claim-relevant model choices, prompts, configurations, evaluation details, and system failure modes at the level needed to support the conclusions being drawn. In OD-ENV-01, that principle is narrower but strict: any hidden system condition that matters to a code must be documented rather than inferred from fluency or apparent personalization.

Several model-integrated-system failure modes can invalidate interpretation. Memory false positive occurs when coders assume stored content was retrieved without evidence that it entered the relevant context. Routing confound occurs when a different model, tool, policy, or context assembly explains an observed change. Version drift occurs when provider updates alter system behavior across episodes. Observer inference occurs when hidden configuration is reconstructed from output style instead of contemporaneous provenance.

Other risks include undocumented summaries influencing later interactions, organizational instructions being mistaken for general model tendencies, permission changes occurring between episodes, and participants retrospectively attributing an action to reliance or authority conflict when the trace shows a different sequence. These risks determine whether a claimed synthetic condition is codable at all. ENV-SYNTH therefore remains a provenance class and never substitutes for identifying the concrete system condition actually supported by the record. The candidate taxonomy should lose access to a label when provenance is insufficient.

16. Human Ecology, Power, and Position

A persistent AI environment is partly defined by who can refuse, inspect, correct, override, leave, or change the system. These capacities are not distributed evenly. A freelancer using a personally chosen assistant, an employee required to use an enterprise agent, and a clinician responsible for an AI-mediated recommendation can encounter the same technical feature under radically different power relations. A generic "user" category would erase exactly the institutional structure the proposed human-side provenance class is supposed to expose.

The Water program therefore treats institutional requirements as more than demographic background. Throughput quotas, liability allocation, management authority, professional ethics, documentation rules, contractual obligations, and labor surveillance can determine whether a nominally user-controlled feature is realistically contestable. A control existing in software is not necessarily equivalent to practical authority in a workplace. The gap between formal and practical authority is therefore a candidate measurement problem rather than a normative assumption.

Future real-thread research creates ethical risks in addition to analytic ones. A workplace transcript may contain confidential information, third-party personal data, client material, internal policy, or content that a worker is not authorized to publish. "Full transcript available" consequently cannot mean "ethically available for study." Initial collection should prioritize low-risk research and project interactions where the participant has authority to provide the material and third-party exposure can be minimized.

17. Research Ethics and Data Protection

Persistent conversation records create unusual research risks because they can accumulate information that was not originally disclosed for research purposes. A single thread may contain third-party names, organizational strategy, legal or health information, financial details, emotional disclosures, credentials, or inferred preferences. Removing a username therefore may not meaningfully de-identify the interaction. The data object itself can be relationally identifying.

Any future corpus must follow data minimization. Only material necessary to preserve the activation event and relevant context should be retained in the analytical record, while raw transcripts should be access-controlled. Publication of examples should use de-identification or synthetic reconstruction unless participants explicitly consent to direct quotation and disclosure is ethically appropriate. Threads containing third-party material require additional caution because primary-participant consent does not automatically authorize disclosure about others.

Institutional confidentiality and employment power also matter. A worker's willingness to provide a transcript does not necessarily establish legal or ethical authority to disclose proprietary or client material. The study should exclude records whose ethical status cannot be established without strain. It should also avoid building an invasive monitoring apparatus in the name of measuring human agency: evidentiary collection should remain proportional to the claim being tested.

18. Candidate Governance Hypotheses

The present paper does not claim that its proposed interventions are established design requirements. They are candidate hypotheses whose failure should be measurable. Different environments may rationally make different tradeoffs among convenience, security, continuity, learning, institutional accountability, and user control. The point is to expose those tradeoffs to measurement.

Failure mode	Candidate intervention	Candidate measurable outcome	What would count against
Stale or mis-scoped memory influences later work	User-visible scope/provenance controls	Correction rate, inappropriate retrieval rate, successful revocation	No improvement or added complexity produces equal/worse control
Confirmation becomes ceremonial	Workflow-valid review design plus institutional metric change	Review duration, provenance inspection, override rate, error detection	Added friction without improved detection or practical authority
Assisted output displaces durable skill	Scaffolding / intermittent unassisted practice	Unassisted performance, persistence, verification ability	Assisted users retain equal skill without added scaffolding
Provider context creates exit cost	Portable/exportable governed state	Successful migration, correction integrity, continuity with provenance	Standard data export performs equally well
Authority hierarchy becomes opaque	Visible source/scope of competing instructions	Correct authority identification, successful contest or reversal	Visibility does not improve authority understanding or remedy

These hypotheses are useful only if tied to a documented failure mode and an observable outcome. An intervention that sounds human-centered but produces no improvement should not be preserved because it aligns with the paper's normative preference. Likewise, some users or institutions may reasonably prefer more automation, greater persistence, or stronger provider-managed context when associated costs are known and accepted. Environmental governance is therefore not synonymous with maximizing user control in every dimension.

19. What This Paper Does Not Claim

The paper does not claim that environments, institutions, technical infrastructures, or social context were absent from previous research. It does not claim that ENV-HUMAN and ENV-SYNTH are validated natural categories. It does not claim that Ecotone is a third state, hidden layer, or universal boundary mechanism. The proposed architecture remains a candidate partition whose technical future depends on adversarial comparison.

The paper does not claim that persistence automatically constitutes Relationship State. It does not treat ordinary memory, retrieval, personalization, or long transcripts as evidence of accumulated relational configuration without

additional support. It does not treat every product defect or unexpected model behavior as a cross-environment phenomenon. Unknown system state is not converted into relationship theory.

The paper also does not claim that AI assistance generally weakens human capability. Existing studies show heterogeneous benefits and harms depending on task, expertise, design, interaction style, and measured outcome. The narrower methodological demand is that assisted output should not substitute for measuring unassisted performance, persistence, verification ability, professional judgment, or other human outcomes when those distinctions matter. Immediate performance and durable capability are separate empirical questions.

Finally, this manuscript does not incorporate the separate resolution-pressure hypothesis concerning unfinished distinctions under abstraction, summarization, convergence, or premature closure. That remains H2 in a separate methods track. Reconstructed demonstrations of conceptual loss are coding stimuli rather than evidence. They are not admissible Water episodes merely because both ideas emerged from the same research conversation.

20. Research Agenda

The immediate priority is not a larger conceptual model. It is the two-thread OD-ENV-01 shakedown under provenance-complete conditions. If the frozen packet cannot produce stable segmentation, comparable records, and meaningful adversarial disagreement, the instrument should be repaired before any larger run. A procedural failure at this stage would be information about the measurement architecture, not evidence against or for the Water distinction itself.

If the shakedown succeeds procedurally, the next step is a preregistered 8-12-thread construct-discrimination run. That run should report segmentation reliability separately from code agreement, preserve AMBIG, compare Water-side outcomes with the prior-art adversary, and publish the disagreement structure rather than only a reconciled final code. No construct revision should occur in the same run report. A run dominated by EXISTING-SUFFICES would remain scientifically interpretable.

Only after a candidate environmental condition survives descriptive discrimination should controlled manipulation begin. Candidate measurable synthetic variables already contained in the research agenda include memory visibility versus hidden memory, confirmation cost measured in time or interaction steps, provenance access open versus closed, correction propagation present versus absent, and override friction low versus high. Candidate human-side manipulations include throughput versus verification framing, explicit accountability assignment, or institutional incentives emphasizing speed versus review. Each manipulation would require a documented baseline, intervention parameter, carryover rule, and outcome rather than an abstract environmental E.

Longer-term research should examine assisted-versus-unassisted performance, verification endurance, correction behavior, authority recognition, portability, and accumulated Relationship State. The objective is not to demonstrate that environments matter. It is to determine whether particular active conditions can be separated well enough to explain who controls them, how they interact, what they change, and whether an established framework already tells us the same thing. The architecture earns continuation only by surviving that competition.

21. Discussion

The most defensible contribution of The Water is not a new theory of context. It is a candidate analytical partition coupled to a refusal to grant that partition technical status merely because it sounds coherent. Persistent AI makes this question worth testing because the synthetic environment can preserve user-specific state, adapt over time, operate through tools, and remain partly opaque while the human remains embedded in institutions with their own histories, authority structures, incentives, and obligations. This combination may or may not require new technical vocabulary.

The conjunction may prove fully reducible to existing sociotechnical language. If a strong STS, human-factors, personalization, or agency account identifies the same evidence, intervention owner, authority structure, governance remedy, measurement requirement, and predicted consequence, Water is conceptual overhead.

The constructed specimen in §7.1 is intentionally such a case: Water can describe it, but it does not beat the rival explanation. A non-confirmatory framework must be able to show examples in which it loses.

If residual value eventually survives, it is likely to be narrow. Certain persistent-AI episodes may become more governable when researchers distinguish the institutional origin of a demand from the engineered mechanism that enforces, translates, remembers, or conflicts with it. The contribution would then be a provenance-sensitive analytical cut rather than a replacement for sociotechnical systems theory. Its value would have to appear in episode-level evidence rather than in the completeness of the manuscript.

22. Conclusion

Persistent human-AI systems can carry histories, permissions, profiles, corrections, and project context across time. Humans entering those systems already inhabit institutions, professions, relationships, markets, cultures, and accountability structures. Some interactions therefore depend on conditions that are not visible in the immediate prompt or output alone. That observation motivates the inquiry but does not establish the proposed taxonomy.

The Water is the metaphor for noticing surrounding configuration. The technical proposal is narrower: a candidate provenance-and-control-surface partition between active inherited human conditions and active engineered synthetic conditions, plus a tightly gated label for observable boundary activation. None of those categories has yet earned empirical validity. The paper therefore ends where the empirical program begins.

Existing frameworks receive first claim on the evidence. If they preserve everything consequential, Water should remain a public-facing metaphor rather than become technical vocabulary. If some residual survives, the residual must be shown through cause, governance, authority, intervention, measurement, or prediction rather than through rhetorical coherence. The point of seeing the water is not to multiply concepts about water; it is to determine whether any backgrounded condition becomes sufficiently observable, governable, and measurable to justify treating it differently.

METHODOLOGICAL SAFEGUARD

Ecotone can fire and Water can still lose.

References

- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775-779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19.
- Ehsan, U., Passi, S., Saha, K., McNutt, T., Riedl, M. O., & Alcorn, S. (2026). From future of work to future of workers: Addressing asymptomatic AI harms to foster dignified human-AI interaction. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, Article 595. <https://doi.org/10.1145/3772318.3791081>
- Fang, X., Xu, W., Ge, Y., Xu, Y., Eckman, S., & Reddy, C. K. (2026). DRIFTLENS: Measuring memory-induced reasoning drift in personalized language models. *arXiv*. <https://doi.org/10.48550/arXiv.2607.02374>
- Gonzalez, C., Donahue, K., Goldstein, D. G., Heidari, H., Jalali, M. S., Schelble, B. G., Singh, A., & Woolley, A. W. (2026). Toward a science of human-AI teaming for decision making: A complementarity framework. *PNAS Nexus*, 5(3), pgag030. <https://doi.org/10.1093/pnasnexus/pgag030>
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392
- Liu, G., Christian, B., Dumbalska, T., Bakker, M. A., & Dubey, R. (2026). AI assistance reduces persistence and hurts independent performance. *arXiv*. <https://doi.org/10.48550/arXiv.2604.04721>
- Navarro, K. F., Syriani, E., & Arawjo, I. (2026). Reporting and reviewing LLM-integrated systems in HCI: Challenges and considerations. *arXiv*. <https://doi.org/10.48550/arXiv.2602.05128>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. <https://doi.org/10.1518/001872097778543886>

Preble, M. C. (2026). Operant Dyad: Defining the task-bound working relationship in human-AI systems. SSRN Working Paper 7234238. <https://doi.org/10.2139/ssrn.7234238>

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676-688. <https://doi.org/10.1016/j.tics.2016.07.002>

Stanford Digital Economy Lab. (2026). The future of personal AI: Portable and persistent personal memory through a Unified Human Context Protocol. Stanford University. <https://digitaleconomy.stanford.edu/publication/the-future-of-personal-ai-portable-and-persistent-personal-memory-through-a-unified-human-context-protocol/>

Zhang, S., Wang, H., & Yi, X. (2025). Exploring collaboration patterns and strategies in human-AI co-creation through the lens of agency: A scoping review of the top-tier HCI literature. Proceedings of the ACM on Human-Computer Interaction, 9(7), Article CSCW413, 1-43. <https://doi.org/10.1145/3757594>

Research Program Status

Object	Current status
The Water	Working metaphor / Perspective
ENV-HUMAN / ENV-SYNTH	Candidate provenance-and-control-surface partition
Ecotone	Candidate activated boundary process; not a state
H1 environmental separation	Held
OD-ENV-01	PRE-PILOT; six-instrument packet frozen
Two-thread shakedown	HELD AT PROVENANCE GATE
8-12 thread run	Not authorized until shakedown holds
H2 resolution pressure	Separate held methods hypothesis

Independent publication. The manuscript is intentionally pre-empirical and should not be represented as a validated taxonomy or as affiliated with Stanford HAI.