

Supplementary Material - Preprint version 1.1

Control Assimilation in Agentic AI: Formal Human Authority Persists as Operational Access to Control Becomes Technically Mediated

Supplementary Tables S1-S6

Contents

- S1. Complete source and manifestation provenance for the 17 frozen architecture cases
- S2. Full-text assessment status at the prespecified analysis freeze
- S3. Case-level control architecture, path, mechanism, and outcome-evidence coding
- S4. Robustness, sensitivity, adversarial stress-testing, and independent reproducibility of the RC configuration framework and outcome evidence
- S5. Multidimensional evidence credibility and source-quality appraisal
- S6. Competing-concept differentiation map for Control Assimilation

Supplementary Table S1

Complete source and manifestation provenance for the 17 frozen architecture cases

Panel A. Source identity and primary bibliographic provenance

Case ID	Work ID	Source role	Architecture / system	Authors	Year	Full title	Venue / repository / source type	DOI / identifier
AP1A001	CAW0019139	Representative primary source	ARMADA real-world multi-robot autonomous rollout + FLOAT failure monitor + human shared control	Wenye Yu; Jun Lv; Zixi Ying; Yang Jin; Chuan Wen; Cewu Lu	2026	ARMADA: Autonomous Online Failure Detection and Human Shared Control Empower Scalable Real-World Deployment and Adaptation	IEEE Robotics and Automation Letters (published work); exact-work arXiv manuscript acquired	10.1109/Ira.2026.3709635
AP1A002	CAW0019672	Representative primary source	AIDR-Cloud agentic production cloud-incident diagnosis/remediation	Akshay Pratinav; Archana Kataria	2026	AIDR-Cloud: An Agentic AI—Driven Incident Response Framework for Cloud Failures	IEEE SoutheastCon 2026	10.1109/southeastcon63549.2026.11476571
AP1A003	CAW0026191	Representative primary source	CTS real-world coordinator multi-agent system under human strategic guidance	Rajiv Maheswaran; Craig Milo Rogers; Romeo Sanchez; Pedro A. Szekely; Gergely Gati; Kevin Smyth; Chris VanBuskirk	2010	Multi-agent systems for the real world	2010 International Symposium on Collaborative Technologies and Systems (CTS)	10.1109/cts.2010.5478452

Case ID	Work ID	Source role	Architecture / system	Authors	Year	Full title	Venue / repository / source type	DOI / identifier
AP1A004	CAW0038849	Representative primary source	Polaris 4 production clinical conversational-AI safety constellation	Subhabrata Mukherjee; Markel Sanz Ausin; Kriti Aggarwal; Debajyoti Datta; Shanil Puri; Woojeong Jin; Tanmay Laud; Neha Manjunath; Jiayuan Ding; Bibek Paudel; Jan Schellenberger; Zepeng Frazier Huo; Walter Shen; Nima Shirazian; Nate Potter; Sathvik Perkari; Darya Filippova; Anton Morozov; Austin Mease; Vivek Muppalla; Ghada Shakir; Alex Miller; Juliana Ghukasyan; Mariska Raglow-Defranco; Maggie Taylor; Herprit Mahal; Jonathan Agnew	2026	Perfecting Human-AI Interaction at Clinical Scale. Turning Production Signals into Safer, More Human Conversations	arXiv / production clinical conversational-AI report	10.48550/arxiv.2603.29893
AP1A005	CAW0040239	Supporting same-work all-versions/concept DOI	AEGIS v1 evidence-driven runtime governance for autonomous agents	Anil Sharma	2026	AEGIS: Evidence-Driven Runtime Governance for Autonomous AI Agents	ANKR Labs technical report / Zenodo	10.5281/zenodo.19998775
AP1A005	CAW0040241	Representative exact concrete version	AEGIS v1 evidence-driven runtime governance for autonomous agents	Anil Sharma	2026	AEGIS: Evidence-Driven Runtime Governance for Autonomous AI Agents	ANKR Labs technical report / Zenodo	10.5281/zenodo.19998776

Case ID	Work ID	Source role	Architecture / system	Authors	Year	Full title	Venue / repository / source type	DOI / identifier
AP1A006	CAW0041092	Representative primary source	Production critical-cloud agent delegated-identity + granular access-control architecture	Arun Malik; Deepal Jayasinghe; Bradley Klemick; Prachi Shah; Nitish Talasu; Vineet Tushar Trivedi	2026	Decentralized Granular Access Control for Agentic AI Systems in Critical Infrastructure	arXiv	10.48550/arxiv.2607.22611
AP1A007	CAW0041584	Representative journal publication	Tiered human-AI FWA quality-management architecture	Venkata Raghavendra Vaishnav Vinjamuri	2026	Human-AI Collaborative Frameworks for Intelligent Fault Detection, Proactive Quality Management, and Augmented Decision Support in Fixed Wireless Access Networks	Journal of Information Systems Engineering and Management / Zenodo versions	10.52783/jisem.v11i3s.15183
AP1A007	CAW0041922	Supporting same-work concrete Zenodo version	Tiered human-AI FWA quality-management architecture	Venkata Raghavendra Vaishnav Vinjamuri	2026	Human-AI Collaborative Frameworks for Intelligent Fault Detection, Proactive Quality Management, and Augmented Decision Support in Fixed Wireless Access Networks	Journal of Information Systems Engineering and Management / Zenodo versions	10.5281/zenodo.21705958
AP1A007	CAW0041929	Supporting same-work all-versions/concept DOI	Tiered human-AI FWA quality-management architecture	Venkata Raghavendra Vaishnav Vinjamuri	2026	Human-AI Collaborative Frameworks for Intelligent Fault Detection, Proactive Quality Management, and Augmented Decision Support in Fixed Wireless Access Networks	Journal of Information Systems Engineering and Management / Zenodo versions	10.5281/zenodo.21705957
AP1A008	CAW0177862	Representative exact Zenodo Version v2	Machine-Native Intelligence v2 208-daemon autonomous system	Rafal Chalupka	2026	Machine-Native Intelligence v2: Operational Evidence from a 208-Daemon Autonomous System	RRC-AI Research Initiative technical report / Zenodo	10.5281/zenodo.19486388 (Version v2); all-versions DOI 10.5281/zenodo.19428196
AP1A009	CAW0203153	Representative primary source	Transparent/controllable multimodal multi-agent recommendation-filtering architecture	Chi Zhang; Zhipeng Xu; Jiahao Liu; Dongsheng Li; Hansu Gu; Peng Zhang; Ning Gu; Tun Lu	2026	Transparent and Controllable Recommendation Filtering via Multimodal Multi-Agent Collaboration	arXiv	10.48550/arxiv.2604.17459

Case ID	Work ID	Source role	Architecture / system	Authors	Year	Full title	Venue / repository / source type	DOI / identifier
AP1A010	CAW0218723	Representative all-versions/concept DOI row	ADAF autonomous multi-cloud infrastructure provisioning and self-healing	Vinay Kumar Reddy Vangoor	2020	Autonomous Infrastructure Provisioning Using AI-Driven DevOps Automation Framework	International Journal of Science, Engineering and Technology / Zenodo archive	10.5281/zenodo.19183280
AP1A010	CAW0218724	Supporting exact concrete Zenodo version	ADAF autonomous multi-cloud infrastructure provisioning and self-healing	Vinay Kumar Reddy Vangoor	2020	Autonomous Infrastructure Provisioning Using AI-Driven DevOps Automation Framework	International Journal of Science, Engineering and Technology / Zenodo archive	10.5281/zenodo.19183279
AP1A011	CAW0224754	Representative primary source	Autonomous production cloud incident-remediation closed loop	Chetan Sasidhar Ravi; Rohit Reddy Patlolla	2024	Autonomous Incident Remediation: A Closed-Loop RAG Framework for Real-Time Root Cause Analysis and Knowledge Synthesis in Distributed Cloud Systems	International Journal of Advanced Research in Science, Communication and Technology	10.48175/ijarsct-19600a
AP1A012	CAW0226114	Representative primary source	AI-driven autonomous offshore power-grid fault detection and self-healing	Ahmad Garba Ahmed	2025	Develop an AI-Driven Fault Detection Model to Autonomously Troubleshoot Electrical Power Grids in High-Risk Offshore Oil Platforms	British Journal of Earth Sciences Research	10.37745/bjesr.2013/vol13n21941
AP1A013	CAW0229550	Representative all-versions/concept DOI row	AI-Driven Release Agents production software-delivery architecture	Sekhar Chittala; Nagaraju Unnava	2026	AI-Driven Release Agents: Leveraging Aiopts And Generative AI Frameworks For Predictive And Reliable Software Delivery	Zenodo repository; same work also published in Journal of International Crisis and Risk Communication Research	10.5281/zenodo.18213125; journal manifestation 10.63278/jicrcr.vi.3565
AP1A013	CAW0229551	Supporting exact concrete Zenodo v1	AI-Driven Release Agents production software-delivery architecture	Sekhar Chittala; Nagaraju Unnava	2026	AI-Driven Release Agents: Leveraging Aiopts And Generative AI Frameworks For Predictive And Reliable Software Delivery	Zenodo repository; same work also published in Journal of International Crisis and Risk Communication Research	10.5281/zenodo.18213126; journal manifestation 10.63278/jicrcr.vi.3565
AP1A014	CAW0229895	Representative all-versions/concept DOI row	CMPSBL OS v6.x.x governed persistent cognitive operating system	Kenneth E. Sweet Jr.	2026	CMPSBL OS Substrate v6.x.x - A Governed Cognitive Operating System for Persistent, Self-Improving AI Systems	CMPSBL technical documentation / Zenodo record package	10.5281/zenodo.18234909
AP1A014	CAW0229907	Supporting exact concrete Zenodo v6.x.x	CMPSBL OS v6.x.x governed persistent cognitive operating system	Kenneth E. Sweet Jr.	2026	CMPSBL OS Substrate v6.x.x - A Governed Cognitive Operating System for Persistent, Self-Improving AI Systems	CMPSBL technical documentation / Zenodo record package	10.5281/zenodo.18425208

Case ID	Work ID	Source role	Architecture / system	Authors	Year	Full title	Venue / repository / source type	DOI / identifier
AP1A015	CAW0232992	Representative all-versions/concept DOI row	Agentic production financial data-pipeline monitoring architecture	Suhas Gorur Ravi Kumar	2026	Agentic Data Pipeline Monitoring: A Multi-Agent Architecture for Anomaly Detection in Financial Transaction Systems	Independent technical report / Zenodo	10.5281/zenodo.20549210
AP1A015	CAW0232993	Supporting exact concrete Zenodo v1	Agentic production financial data-pipeline monitoring architecture	Suhas Gorur Ravi Kumar	2026	Agentic Data Pipeline Monitoring: A Multi-Agent Architecture for Anomaly Detection in Financial Transaction Systems	Independent technical report / Zenodo	10.5281/zenodo.20549211
AP1A016	CAW0233812	Representative primary source	Deployed RL/IoT human-centric energy-intervention architecture	Rafael Neves Teixeira	2026	Gamifying Sustainability: Harnessing AI and IoT for Human-Centric Energy Efficiency	University of Porto, FEUP, MSc thesis	
AP1A017	CAW0242392	Representative exact concrete Zenodo v1	PulseADT deployed agentic-security defense-transformer architecture	David Idris	2026	Autonomous Defense Transformers: Security-Native Reasoning for Digital Infrastructure	Glemad technical paper / Zenodo	10.5281/zenodo.18878123
AP1A017	CAW0242394	Supporting all-versions/concept DOI	PulseADT deployed agentic-security defense-transformer architecture	David Idris	2026	Autonomous Defense Transformers: Security-Native Reasoning for Digital Infrastructure	Glemad technical paper / Zenodo	10.5281/zenodo.18878122

Panel B. Manifestation identity, evidence tier, and provenance resolution

Case ID	PMID	Stable URL	Evidence tier	Manifestation / version relationship	Identity status	Identity / provenance note	Metadata basis
AP1A001	Not identified in canonical metadata	https://doi.org/10.1109/lra.2026.3709635	E1_DIRECT	Representative primary source	RESOLVED	Frozen publication DOI is IEEE RA-L 10.1109/LRA.2026.3709635. Acquired exact-work manuscript is an arXiv v1 manifestation cross-linked by the authors' project page; not labelled IEEE Version of Record.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A002	Not identified in canonical metadata	https://doi.org/10.1109/southeastcon63549.2026.11476571	E1_DIRECT	Representative primary source	RESOLVED	Exact IEEE proceedings manifestation established; authoritative exact full text was subsequently acquired and identity-reconciled. The later acquisition did not alter the prespecified analysis freeze or transfer substantive coding.	Canonical P1 checkpoint + Phase-1 frozen coding

Case ID	PMID	Stable URL	Evidence tier	Manifestation / version relationship	Identity status	Identity / provenance note	Metadata basis
AP1A003	Not identified in canonical metadata	https://doi.org/10.1109/cts.2010.5478452	E1_DIRECT	Representative primary source	RESOLVED	Exact queued DOI/title and author-uploaded CTS 2010 full text matched; no neighboring AAMAS manifestation used for coding.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A004	Not identified in canonical metadata	https://arxiv.org/abs/2603.29893	E1_DIRECT	Representative primary source	RESOLVED	Exact queued title/DOI/full-text manifestation matched.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A005	Not identified in canonical metadata	https://zenodo.org/records/19998775	E1_DIRECT	Supporting same-work all-versions/concept DOI	RESOLVED	Provider-declared all-versions/concept DOI; same work as concrete record CAW0040241.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A005	Not identified in canonical metadata	https://zenodo.org/records/19998776	E1_DIRECT	Representative exact concrete version	RESOLVED	Concrete Zenodo version DOI; exact AEGIS v1 artifact acquired.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A006	Not identified in canonical metadata	https://arxiv.org/abs/2607.22611	E1_DIRECT	Representative primary source	RESOLVED	Exact queued arXiv DOI 10.48550/arxiv.2607.22611 matched; complete exact arXiv HTML used.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A007	Not identified in canonical metadata	https://doi.org/10.52783/jisem.v11i3s.15183	E1_DIRECT	Representative journal publication	RESOLVED	Journal publication used as representative source.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A007	Not identified in canonical metadata	https://zenodo.org/records/21705958	E1_DIRECT	Supporting same-work concrete Zenodo version	RESOLVED	Concrete Zenodo version of the same FWA work.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A007	Not identified in canonical metadata	https://zenodo.org/records/21705957	E1_DIRECT	Supporting same-work all-versions/concept DOI	RESOLVED	All-versions/concept Zenodo DOI for the same FWA work.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A008	Not identified in canonical metadata	https://zenodo.org/records/19486388	E1_DIRECT	Representative exact Zenodo Version v2	RESOLVED_VERSION_ROLE	Identity reconciled: canonical DOI 10.5281/zenodo.19486388 is exact Version v2; provider also identifies 10.5281/zenodo.19428196 as the all-versions DOI. Both roles are preserved; no conflict remains.	Canonical P1 checkpoint + Phase-1 frozen coding + Manual Acquisition Ledger/provider version evidence

Case ID	PMID	Stable URL	Evidence tier	Manifestation / version relationship	Identity status	Identity / provenance note	Metadata basis
AP1A009	Not identified in canonical metadata	https://arxiv.org/abs/2604.17459	E1_E2_MIXED	Representative primary source	RESOLVED	Exact arXiv DOI/title match.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A010	Not identified in canonical metadata	https://zenodo.org/records/19183280	E1_DIRECT	Representative all-versions/concept DOI row	RESOLVED	Provider/frozen row represents the all-versions/concept DOI; exact-work artifact shared with concrete version row.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A010	Not identified in canonical metadata	https://zenodo.org/records/19183279	E1_DIRECT	Supporting exact concrete Zenodo version	RESOLVED	Concrete Zenodo version; same work as CAW0218723.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A011	Not identified in canonical metadata	https://doi.org/10.48175/ijarsct-19600a	E1_DIRECT	Representative primary source	RESOLVED	Exact DOI/title/article manifestation confirmed.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A012	Not identified in canonical metadata	https://doi.org/10.37745/bjesr.2013/vol13n21941	E1_DIRECT	Representative primary source	RESOLVED	Exact DOI/title/article manifestation confirmed.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A013	Not identified in canonical metadata	https://zenodo.org/records/18213125	E2_PRIMARY_TECHNICAL	Representative all-versions/concept DOI row	RESOLVED	Frozen Zenodo concept DOI. Same-title journal publication exists at DOI 10.63278/jicrcr.vi.3565; journal DOI is recorded as publication-chain metadata, not substituted for the frozen Zenodo row.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A013	Not identified in canonical metadata	https://zenodo.org/records/18213126	E2_PRIMARY_TECHNICAL	Supporting exact concrete Zenodo v1	RESOLVED	Concrete Zenodo v1; same work as CAW0229550. Journal manifestation DOI 10.63278/jicrcr.vi.3565 recorded separately.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A014	Not identified in canonical metadata	https://zenodo.org/records/18234909	E1_DIRECT	Representative all-versions/concept DOI row	RESOLVED	All-versions/concept DOI. Authoritative provider evidence binds this concept row to concrete Version v6.x.x CAW0229907 despite non-adjacent identifier numbering.	Canonical P1 checkpoint + Phase-1 frozen coding + Manual Acquisition Ledger/provider version evidence
AP1A014	Not identified in canonical metadata	https://zenodo.org/records/18425208	E1_DIRECT	Supporting exact concrete Zenodo v6.x.x	RESOLVED	Exact concrete Version v6.x.x; provider-declared member of concept DOI 10.5281/zenodo.18234909 family.	Canonical P1 checkpoint + Phase-1 frozen coding + Manual Acquisition Ledger/provider version evidence

Case ID	PMID	Stable URL	Evidence tier	Manifestation / version relationship	Identity status	Identity / provenance note	Metadata basis
AP1A015	Not identified in canonical metadata	https://zenodo.org/records/20549210	E1_E2_MIXED	Representative all-versions/concept DOI row	RESOLVED	All-versions/concept DOI; exact-work artifact shared with concrete v1 row.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A015	Not identified in canonical metadata	https://zenodo.org/records/20549211	E1_E2_MIXED	Supporting exact concrete Zenodo v1	RESOLVED	Concrete Zenodo v1; exact PDF acquired.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A016	Not identified in canonical metadata	https://sigarra.up.pt/feup/pt/teses.lista_teses?p_curso=22862&p_lang=1	E2_PRIMARY_TECHNICAL	Representative primary source	RESOLVED	No DOI in canonical metadata. Exact FEUP MSc thesis acquired; stable institutional thesis-list source retained.	Canonical P1 checkpoint + Phase-1 frozen coding
AP1A017	Not identified in canonical metadata	https://zenodo.org/records/18878123	E2_PRIMARY_TECHNICAL	Representative exact concrete Zenodo v1	RESOLVED	Exact concrete Zenodo v1; provider identifies CAW0242394 DOI as all-versions/concept DOI.	Canonical P1 checkpoint + Phase-1 frozen coding + Manual Acquisition Ledger/provider version evidence
AP1A017	Not identified in canonical metadata	https://zenodo.org/records/18878122	E2_PRIMARY_TECHNICAL	Supporting all-versions/concept DOI	RESOLVED	All-versions/concept DOI resolving to the concrete work family; exact v1 file acquired from provider family.	Canonical P1 checkpoint + Phase-1 frozen coding + Manual Acquisition Ledger/provider version evidence

Supplementary Table S2

Full-text assessment status at the prespecified analysis freeze

Status category	n	Proportion of P1 (n=1,156)	Substantive FT assessment completed?	Meaning at freeze	Permitted manuscript wording	Not permitted	Provenance note
Substantively full-text assessed	990	0.856401384083045	Yes	Included in empirical architecture-eligibility review.	Substantively full-text assessed / empirically architecture-assessed.	Do not imply all 990 were retained architecture cases.	753 near_match_FT + 237 not_near_match_FT; novelty label did not determine architecture eligibility.
Not substantively full-text assessed at analysis freeze	166	0.14359861591695502	No	Frozen aggregate of records whose substantive full-text adjudication was not completed at the prespecified analysis freeze.	166 records were not substantively full-text assessed at the prespecified analysis freeze.	Do not treat these records as substantively excluded or empirically ineligible.	Submission-facing table reports the frozen substantive-assessment state only. Operational acquisition and identity workflow subcategories are not reproduced because they are not substantive eligibility categories; the frozen denominator is unchanged.
P1 total	1156	1	Mixed	Priority screening population.	P1 priority set, n=1,156.	Do not use as final architecture denominator.	990 + 166 = 1,156.
Exploratory attrition diagnostic by prespecified P1 priority stratum							
P1 stratum	Total	Substantively assessed	Not assessed	Non-adjudication proportion	Diagnostic contribution	Interpretation	Guardrail
P1A	172	150	22	0.12790697674418605	First-wave highest-priority stratum	12.8% non-adjudicated	Workflow balance only
P1B	85	70	15	0.17647058823529413	Intermediate priority	17.6% non-adjudicated	Workflow balance only
P1C	899	770	129	0.14349276974416017	Broad first-wave priority stratum	14.3% non-adjudicated	Workflow balance only
Total	1156	990	166	0.14359861591695502	Pearson $\chi^2(2)=1.09$; p=0.579; Cramér's V=0.031	No detectable concentration across P1A/B/C	Does not establish MCAR/MAR or absence of attrition bias
Diagnostic interpretation: substantive FT non-adjudication was not detectably concentrated in a prespecified P1 priority stratum (P1A 12.8%, P1B 17.6%, P1C 14.3%; $\chi^2(2)=1.09$, p=0.579; Cramér's V=0.031). This test addresses workflow balance only. It is not evidence that the 166 records are missing at random, unbiased, or substantively irrelevant.							

Supplementary Table S3

Case-level control architecture, path, mechanism, and outcome-evidence coding

Panel A. Architecture, configuration, deployment, evidence, and pre-action opportunity

Case ID	Architecture / system	RC	Membership	Stratum	Deployment	Outcome evaluable	Evidence tier	Pre-action opportunity
AP1A001	ARMADA multi-robot shared control	RC-B	HIGH	Embodied	D2	yes	E1_DIRECT	NO_CLEAR
AP1A002	AIDR-Cloud incident response	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A003	CTS real-world coordinator MAS	RC-B	MODERATE	Digital	D2	yes	E1_DIRECT	NO_CLEAR
AP1A004	Polaris clinical safety constellation	RC-B	HIGH	Digital	D1	yes	E1_DIRECT	NO_CLEAR
AP1A005	AEGIS runtime governance	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A006	Critical-cloud delegated identity / access control	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A007	Tiered human-AI FWA quality management	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A008	Machine-Native Intelligence v2	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A009	Transparent/controllable recommendation filtering	RC-C	HIGH	Digital	D1	yes	E1_E2_MIXED	YES
AP1A010	ADAF multi-cloud provisioning / self-healing	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A011	Autonomous cloud incident remediation	RC-A	HIGH	Digital	D1	yes	E1_DIRECT	YES
AP1A012	Autonomous offshore power-grid fault management	RC-B	MODERATE	Embodied	D1	no	E1_DIRECT	UNKNOWN_TIMELINE SS
AP1A013	AI-Driven Release Agents	RC-A	HIGH	Digital	D1	yes	E2_PRIMARY_TECHNICAL	YES
AP1A014	CMPSBL OS v6.x.x	RC-A	HIGH	Digital	D1	no	E1_DIRECT	YES
AP1A015	Agentic financial data-pipeline monitoring	RC-A	HIGH	Digital	D1	yes	E1_E2_MIXED	YES
AP1A016	RL/IoT human-centric energy intervention	RC-C	HIGH	Digital	D1	yes	E2_PRIMARY_TECHNICAL	YES
AP1A017	PulseADT autonomous defense transformer	RC-A	HIGH	Digital	D1	no	E2_PRIMARY_TECHNICAL	YES

Panel B. Local control mechanisms, independent constraints, path structure, and attention recruitment

Case ID	Stop	Veto	Override / takeover	Rollback	Recovery	ICC Variant B	Role-path count	MMCD path minima / qualifiers	Attention-recruitment coding
AP1A001	UNKNOWN	UNKNOWN	SUPPORTED	SUPPORTED	SUPPORTED	UNKNOWN	1	3 (MINIMUM)	CLEAR
AP1A002	UNKNOWN	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	1	6 (MINIMUM)	CLEAR
AP1A003	UNKNOWN	UNKNOWN	PARTIAL	UNKNOWN	PARTIAL	UNKNOWN	2	1 (MINIMUM); 0 (documented minimum, not true system depth)	INDETERMINATE
AP1A004	UNKNOWN	UNKNOWN	SUPPORTED	UNKNOWN	SUPPORTED	UNKNOWN	2	4 (MINIMUM); 1 path UNKNOWN_NUMERIC	CLEAR
AP1A005	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	3	2; 1; 0 documented minimum (not true system depth)	CLEAR
AP1A006	UNKNOWN	SUPPORTED	PARTIAL	UNKNOWN	UNKNOWN	SUPPORTED	3	2; 1; DIRECT_PARALLEL_PATH	CLEAR
AP1A007	UNKNOWN	SUPPORTED	SUPPORTED	PARTIAL	PARTIAL	UNKNOWN	2	3; 3 (AT_LEAST)	CLEAR
AP1A008	UNKNOWN	SUPPORTED	UNKNOWN	UNKNOWN	QUALITATIVE	SUPPORTED_SCOPE_LIMITED	1	3 (AT_LEAST)	CLEAR
AP1A009	UNKNOWN	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	UNKNOWN	3	1; 3; DIRECT_HUMAN_INPUT	MIXED
AP1A010	UNKNOWN	PARTIAL	UNKNOWN	UNKNOWN	SUPPORTED	SUPPORTED_SCOPE_LIMITED	2	3; 4 (AT_LEAST)	CLEAR
AP1A011	UNKNOWN	PARTIAL	PARTIAL	UNKNOWN	QUALITATIVE	SUPPORTED_SCOPE_LIMITED	2	2; 4 (AT_LEAST)	CLEAR
AP1A012	UNKNOWN	UNKNOWN	PARTIAL	UNKNOWN	SUPPORTED	UNKNOWN	1	3 (AT_LEAST)	INDETERMINATE
AP1A013	UNKNOWN	UNKNOWN	PARTIAL	SUPPORTED	PARTIAL	UNKNOWN	2	3; 3 (AT_LEAST)	CLEAR
AP1A014	PARTIAL	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED	SUPPORTED_SCOPE_LIMITED	2	2 (AT_LEAST); 1 path UNKNOWN_NUMERIC	CLEAR
AP1A015	UNKNOWN	SUPPORTED	UNKNOWN	UNKNOWN	QUALITATIVE	UNKNOWN	1	4 (AT_LEAST)	CLEAR
AP1A016	UNKNOWN	SUPPORTED	SUPPORTED	UNKNOWN	UNKNOWN	UNKNOWN	1	3 (AT_LEAST)	CLEAR

Case ID	Stop	Veto	Override / takeover	Rollback	Recovery	ICC Variant B	Role-path count	MMCD path minima / qualifiers	Attention-recruitment coding
AP1A017	UNKNOWN	SUPPORTED	UNKNOWN	SUPPORTED	SUPPORTED	SUPPORTED_SCOPE_LIMITED	2	4; 2 (AT_LEAST)	CLEAR

Panel C. Attention routing, outcome evidence, observed human control, overlay, and limitations

Case ID	Attention-routing mechanism	Outcome-evidence class	Outcome-evidence confidence	Observed human act / control path	O1 overlay	Key case-level limitation	Primary reference
AP1A001	failure detection / routing	DIRECT_HUMAN_EXERCISE_OBSERVED	HIGH	Teleoperator takeover/correction after FLOAT failure routing	No	No clear universal pre-action gate; timely access depends on machine failure detection.	Yu et al., 2026
AP1A002	risk routing	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	High-risk approval before remediation	No	Completed human approvals not separately quantified.	Pratinav & Kataria, 2026
AP1A003	human strategy / monitoring unclear	INDIRECT_OR_MACHINE_ONLY_CONTROL_RELLEVANCE	HIGH	Strategic guidance plus technical fallback	No	Trigger/timing for runtime takeover incompletely specified; hybrid ambiguity retained.	Maheswaran et al., 2010
AP1A004	machine-selected escalation / transfer	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	Transfer/review after machine routing	No	Completed human review contribution not quantified.	Mukherjee et al., 2026
AP1A005	policy GATE routing	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	GATE approval token; promotion veto; global/service rollback	Yes	Strong governance architecture does not isolate causal human contribution.	Sharma, 2026
AP1A006	authorization / risk gate	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	High-risk multi-party approval; breakglass path	No	Breakglass semantics and human-specific outcome denominator incomplete.	Malik et al., 2026
AP1A007	tier routing	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	Tier-2/Tier-3 approval/review	No	Universal rollback and response latency not established.	Vinjamuri, 2026
AP1A008	quarantine / class routing	DIRECT_HUMAN_EXERCISE_OBSERVED	MODERATE	Manual Level-D verification / epistemic promotion gate	Yes	Direct human exercise is narrow and test-specific.	Chalupka, 2026

Case ID	Attention-routing mechanism	Outcome-evidence class	Outcome-evidence confidence	Observed human act / control path	O1 overlay	Key case-level limitation	Primary reference
AP1A009	user initiation + machine-mediated filtering	DIRECT_HUMAN_EXERCISE_OBSERVED	HIGH	Preview, appeal/dispute, preference reweighting	No	RC-C depends on explicitly locked end-user/final-action role precedence.	Zhang et al., 2026
AP1A010	policy-exception routing	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	Human exception handling outside bounded policy envelope	No	Fine-grained human intervention mechanics remain partial.	Vangoor, 2020
AP1A011	severity / confidence routing	CONTROL_EFFECT_OBSERVED_HUMAN_CONTRIBUTION_NOT_ISOLATED	HIGH	Mandatory P0 and low-confidence human handling	No	Completed human review/intervention not denominatorized.	Ravi & Patlolla, 2024
AP1A012	residual dashboard / override unclear	NOT_OUTCOME_EVALUABLE	—	Residual operator dashboard / override	No	Timely opportunity and narrow effective-control outcome not established.	Ahmed, 2025
AP1A013	risk / uncertainty routing	INDIRECT_OR_MACHINE_ONLY_CONTROL_RELLEVANCE	HIGH	Selected high-risk/critical review; automatic halt/rollback	No	Observed recovery is primarily machine-mediated rather than exercised human control.	Chittala & Unnava, 2026
AP1A014	riskied-change gate	NOT_OUTCOME_EVALUABLE	—	Pre-execution CORTEX/high-risk approval; circuit/abort controls	Yes	Strong architecture but no narrow observed downstream human-control consequence.	Sweet Jr., 2026
AP1A015	severity / whitelist routing	DIRECT_HUMAN_EXERCISE_OBSERVED	MODERATE	Owner confirmation / active investigation after drift detection	No	Architecture-level attribution of reported response-time improvement to human action is limited.	Ravi Kumar, 2026
AP1A016	attention selection / suppression	DIRECT_HUMAN_EXERCISE_OBSERVED	HIGH	Accept/reject/ignore/no-response; final physical action	No	n=1 system; final human action remains upstream machine-mediated.	Teixeira, 2026
AP1A017	action-class routing	NOT_OUTCOME_EVALUABLE	—	Class-3 approval; policy/model-change approval	Yes	Operational performance reported without narrow observed human-control consequence.	Idris, 2026

Supplementary Table S4

Robustness, sensitivity, adversarial stress-testing, and independent reproducibility of the RC configuration framework and outcome evidence

A. Configuration sensitivity scenarios

Scenario	Population / rule	n	RC-A	RC-B	RC-C	Set aside / moved	What changes?	Interpretation	Inference boundary
S0 — reference	All frozen architectures; locked role-priority decision tree	17	11	4	2	None	Reference distribution	Taxonomy exhaustive across frozen population.	Descriptive sample only.
S1 — outcome evaluable	Outcome-evaluable cases only	14	9	3	2	A012, A014, A017 excluded from outcome layer	Structure persists.	Supports using RCs descriptively within outcome subset.	Not an effectiveness comparison.
S2 — Digital + D1 outcome subset	Outcome-evaluable and Digital + D1 only	12	9	1	2	A001, A003 removed	RC-B shrinks 3 → 1.	RC-B is materially deployment/stratum sensitive.	No independent configuration effect.
S3 — E1_DIRECT restriction	Outcome-evaluable + E1_DIRECT evidence only	10	7	3	0	RC-C cases absent	RC-C disappears.	RC-C is a qualitative boundary pattern, not robust prevalence class across evidence tiers.	Do not generalize RC-C frequency.
S4 — HIGH membership only	Outcome-evaluable; retain only HIGH RC membership	13	9	2	2	A003 set aside	RC-A/RC-C stable; RC-B shrinks.	Main architecture direction persists without moderate A003.	A012 already outside outcome subset.
S5 — gate-first counterfactual	Outcome-evaluable; classify generic pre-action gate before end-user role	14	11	3	0	A009 & A016 move RC-C → RC-A	RC-C collapses.	Shows explicit dependence on the role/final-action precedence rule.	This is a taxonomy-rule sensitivity, not hidden miscoding.

Mechanism-state sensitivity: supported-only versus supported + partial

Configuration	Mechanism	Supported only	Supported + partial	n	Direction	Interpretation	Counterweight	Allowed claim	Blocked claim
RC-A	Veto	6	8	9	Persists	Gate/veto signature remains dominant under partial-inclusive coding.	One RC-A case remains unknown for veto.	Selective gating is robustly documented in RC-A.	No effectiveness superiority.
RC-A	Override / takeover	3	6	9	Variable	Override is secondary and more coding-sensitive than veto.	Three remain unknown.	Override capability is heterogeneous within RC-A.	Do not define RC-A by override.

Configurat ion	Mechanism	Supported only	Supported + partial	n	Direction	Interpretation	Counterweight	Allowed claim	Blocked claim
RC-B	Override / takeover	2	3	3	Persists	Takeover/override signature persists with or without partial inclusion.	n=3 outcome cases.	RC-B is descriptively takeover-oriented.	No generalization beyond small subset.
RC-C	Veto	2	2	2	Unchanged	Both end-user cases retain reject/veto capability.	n=2; evidence tiers differ.	Terminal human agency includes rejection capability in both cases.	No superiority.
All outcome	Global/system stop	1	1	14	Unchanged	Global stop remains rare under either threshold.	Unknown does not equal absent.	Documented stop is much less common than local mechanisms.	Do not claim absence in unknown cases.

ICC and direct-human-exercise sensitivity

Test	Reference coding	Stricter coding	RC-A	RC-B	RC-C	Cases affected	Result	Interpretation	Guardrail
ICC	SUPPORTED + SUPPORTED_SCOPE_ LIMITED	SUPPORTED only	6/9 → 3/9	0/3 → 0/3	0/2 → 0/2	RC-A scope- limited cases	Direction unchanged; magnitude decreases.	Only RC-A contains documented Variant-B ICC in outcome subset, but taxonomy and ICC are structurally entangled.	No independent association test.
Direct human exercise	Base: HIGH + MODERATE direct	Strict: HIGH direct only	2/9 → 0/9	1/3 → 1/3	2/2 → 2/2	A008, A015 recoded to non-isolated	Total direct 5/14 → 3/14.	Main conclusion strengthens: direct human exercise is less robustly demonstrated than architecture-level authority.	Do not treat RC-C n=2 as superior.

Adversarial membership stress-test

Case	Frozen RC	Membersh ip	Strongest alternative	Attack	Reason frozen RC retained	Residual uncertainty	Rule sensitivity	Final verdict	Freeze implication
A003	RC-B	MODERAT E	Hybrid / unclassified	Strategic human guidance may not fit either gate or exception logic cleanly.	No clear runtime pre-action gate; monitoring intervention is closest to exception/fallback.	Trigger/timing incompletely specified.	No	Retain RC-B MODER ATE	Uncertainty must remain visible.
A012	RC-B	MODERAT E	Residual authority / unclassified	Dashboard + override may overstate timely takeover capability.	Residual override documented; universal timely pre-action gate not documented.	Timeliness and trigger unclear; outcome not evaluable.	No	Retain RC-B MODER ATE	Do not infer timely takeover.
A009	RC-C	HIGH	RC-A	Preview and appeal can be represented as gates.	Responsible human is end user retaining terminal accept/reject/appeal after upstream filtering.	n=1 RC-C architecture in this mechanism family.	YES	Retain RC-C	Role/final-action precedence locked.

Case	Frozen RC	Membersh ip	Strongest alternative	Attack	Reason frozen RC retained	Residual uncertainty	Rule sensitivity	Final verdict	Freeze implication
A016	RC-C	HIGH	RC-A	Presented nudge rejection is a pre-effect gate.	End user retains final physical action after machine selection/suppression.	E2 technical evidence; n=1 architecture.	YES	Retain RC-C	Role/final-action precedence locked.

Frozen interpretation boundary: RC-A/B/C are descriptive recurring configurations, not latent natural classes, an autonomy scale, or a causal exposure. A003/A012 retain localized classification uncertainty. RC-C depends explicitly on end-user/final-action role precedence. Cross-stratum pooling, configuration superiority, prevalence extrapolation, OCR-rate comparison, causal claims, and any composite Control Assimilation score remain blocked.

Independent blinded RC reclassification — pre-adjudication reproducibility test

Design	Material	Blinding	n	Exact RC agreement	Cohen's kappa	Confidence agreement	Adjudication n	Reviewer blinding / process	Interpretive boundary
Independent reclassification after taxonomy freeze	Standardized neutral architecture summaries + locked RC decision tree	Reviewer did not receive original AP1A case IDs, original RC labels, outcome classifications, manuscript, or source papers	17	17/17 (100%)	1.000	16/17 (94.1%)	None required for RC labels	Returned file frozen before comparison; no RC adjudication required	Demonstrates reproducibility of RC labels from standardized summaries; does not equal independent primary-source extraction or independent taxonomy derivation.

Case-level independent reclassification

Case	Original RC	Independe nt RC	RC agreeme nt	Original confidence	Independent confidence	Confidence agreement	Adjudication required?	Reviewer-process note	Interpretation
A001	RC-B	RC-B	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A002	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A003	RC-B	RC-B	YES	MODERATE	MODERATE	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A004	RC-B	RC-B	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A005	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication

Case	Original RC	Independent RC	RC agreement	Original confidence	Independent confidence	Confidence agreement	Adjudication required?	Reviewer-process note	Interpretation
A006	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A007	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A008	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A009	RC-C	RC-C	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A010	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A011	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A012	RC-B	RC-B	YES	MODERATE	LOW	NO	NO	Independent blinded reclassification; no adjudication required	Exact RC replication; independent confidence lower
A013	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A014	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A015	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A016	RC-C	RC-C	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication
A017	RC-A	RC-A	YES	HIGH	HIGH	YES	NO	Independent blinded reclassification; no adjudication required	Exact RC replication

Pre-adjudication rule: the returned reviewer workbook was frozen before comparison. Because all 17 RC labels agreed exactly, no RC adjudication was performed or required. The only confidence disagreement was A012 (original MODERATE; independent LOW), which reinforces the pre-existing uncertainty around trigger/timeliness without changing the frozen RC-B membership. The independent exercise therefore supports rule-level reproducibility of the three-category RC classification, but not independent extraction from primary sources.

Supplementary Table S5

Multidimensional evidence credibility and source-quality appraisal

Panel A. Publication status, evidence directness, deployment, mechanism specificity, and outcome attribution

Case	Representative source	Publication form	Formal publication channel	Evidence directness	Deployment evidence	Human-control mechanism specificity	Outcome attribution
A001	Yu et al., 2026; IEEE RA-L	Peer-reviewed journal article	YES	E1_DIRECT	D2 real-world rollout	HIGH	Direct human exercise, HIGH
A002	Pratinav & Kataria, 2026; IEEE SoutheastCon	Conference proceeding	YES	E1_DIRECT	D1	HIGH	Control effect; human contribution not isolated
A003	Maheswaran et al., 2010; CTS	Conference proceeding	YES	E1_DIRECT	D2 field system	MODERATE	Indirect/machine-control relevance
A004	Mukherjee et al., 2026; arXiv	Preprint	NO	E1_DIRECT	D1 production	HIGH	Control effect; human contribution not isolated
A005	Sharma, 2026; ANKR Labs/Zenodo	Technical report / repository	NO	E1_DIRECT	D1	HIGH	Control effect; human contribution not isolated
A006	Malik et al., 2026; arXiv	Preprint	NO	E1_DIRECT	D1	HIGH	Control effect; human contribution not isolated
A007	Vinjamuri, 2026; JISEM	Journal article + repository versions	YES	E1_DIRECT	D1	HIGH	Control effect; human contribution not isolated
A008	Chalupka, 2026; RRC-AI/Zenodo	Technical report / repository	NO	E1_DIRECT	D1	HIGH	Direct human exercise, MODERATE
A009	Zhang et al., 2026; arXiv	Preprint	NO	E1_E2_MIXED	D1	HIGH	Direct human exercise, HIGH
A010	Vangoor, 2020; IJSET/Zenodo	Journal article / repository archive	YES	E1_DIRECT	D1	MODERATE-HIGH	Control effect; human contribution not isolated
A011	Ravi & Patlolla, 2024; IJARST	Journal article	YES	E1_DIRECT	D1	HIGH	Control effect; human contribution not isolated
A012	Ahmed, 2025; BJESR	Journal article	YES	E1_DIRECT	D1 embodied/critical infrastructure	MODERATE	Not outcome-evaluable
A013	Chittala & Unnava, 2026; JICRCR + Zenodo	Journal article + repository manifestation	YES	E2_PRIMARY_TECHNICAL	D1	MODERATE-HIGH	Indirect/machine-control relevance

Case	Representative source	Publication form	Formal publication channel	Evidence directness	Deployment evidence	Human-control mechanism specificity	Outcome attribution
A014	Sweet Jr., 2026; CMPSBL/Zenodo	Technical documentation / repository	NO	E1_DIRECT	D1	HIGH	Not outcome-evaluable
A015	Ravi Kumar, 2026; Zenodo	Independent technical report	NO	E1_E2_MIXED	D1	MODERATE-HIGH	Direct human exercise, MODERATE
A016	Teixeira, 2026; FEUP MSc thesis	Academic thesis	NO_JOURNAL_P EER_REVIEW	E2_PRIMARY _TECHNICAL	D1	HIGH	Direct human exercise, HIGH
A017	Idris, 2026; Glemad/Zenodo	Technical paper / repository	NO	E2_PRIMARY _TECHNICAL	D1	HIGH	Not outcome-evaluable

Panel B. Validation, self-report dependence, traceability, credibility limitations, and interpretive use

Case	Independent external validation	Developer/ author self-report dependence	Exact-source traceability	Credibility strengths	Primary credibility limitation	Formal-publication sensitivity set	RC	Interpretive use
A001	No independent replication identified	MODERATE	HIGH	Formal journal publication; real-world deployment; exercised takeover	Single research program; outcome attribution not randomized	IN	RC-B	High-credibility architecture exemplar
A002	No independent validation identified	MODERATE-HIGH	HIGH	Formal conference publication; explicit risk-gated architecture	Human approval numerator/outcome contribution not isolated	IN	RC-A	Architecture evidence, not human-effect estimate
A003	No independent replication identified	MODERATE	HIGH	Formal conference publication; real-world coordination	Human takeover trigger/timing incompletely specified	IN	RC-B	Boundary case; membership MODERATE
A004	No peer-reviewed version identified in frozen provenance	HIGH	HIGH	Large-scale production signals; detailed escalation architecture	Preprint; developer/author-reported production outcomes	OUT	RC-B	Mechanism evidence with publication-status caveat
A005	No independent external validation identified	HIGH	HIGH	Executable gates, kill/rollback paths, exact artifact provenance	Technical self-report; no independent replication	OUT	RC-A	Strong architecture mechanics; outcome attribution limited
A006	No peer-reviewed version identified	HIGH	HIGH	Detailed delegated-identity and authorization mechanisms	Preprint and self-reported operational metrics	OUT	RC-A	Mechanism-level evidence
A007	Independent replication not identified	HIGH	HIGH	Formal publication channel; explicit tiered human approval	Venue/review rigor not independently audited; self-reported outcomes	IN	RC-A	Retained with credibility caveat

Case	Independent external validation	Developer/ author self-report dependence	Exact-source traceability	Credibility strengths	Primary credibility limitation	Formal-publication sensitivity set	RC	Interpretive use
A008	No independent validation identified	HIGH	HIGH	Concrete v2 artifact; exact version identity resolved; explicit quarantine gate	Technical self-report; direct exercise narrow/test-specific	OUT	RC-A	Narrow direct-control exemplar
A009	No peer-reviewed version identified	HIGH	HIGH	Explicit user preview/appeal/reweight pathways	Preprint; RC-C evidence-tier sensitivity	OUT	RC-C	Qualitative terminal-agency exemplar
A010	Independent replication not identified	HIGH	HIGH	Formal publication channel; bounded policy-envelope mechanics	Publication/review rigor not independently audited; human mechanics partial	IN	RC-A	Architecture evidence with caution
A011	Independent replication not identified	HIGH	HIGH	Formal publication channel; explicit severity/confidence routing	Human review numerator and causal contribution not isolated	IN	RC-A	Strong routing example
A012	Independent validation not identified	HIGH	HIGH	Formal publication channel; embodied critical-infrastructure architecture	Timeliness/control consequence not established; membership MODERATE	IN	RC-B	Negative/boundary evidence
A013	Independent replication not identified	HIGH	HIGH	Formal publication chain; explicit risk/rollback architecture	Operational recovery primarily machine-mediated; source quality not independently audited	IN	RC-A	Architecture/recovery evidence
A014	No independent validation identified	HIGH	HIGH	Detailed governed-change and recovery controls; exact version mapping	Developer technical documentation; no narrow human-control outcome	OUT	RC-A	Governance-architecture exemplar
A015	No independent validation identified	HIGH	HIGH	Exact technical artifact; owner-confirmed drift response	Self-report; causal human attribution limited	OUT	RC-A	Moderate direct-control evidence
A016	Academic examination; no independent replication identified	MODERATE-HIGH	HIGH	Institutional thesis; explicit final user action and measured engagement	Not journal peer-reviewed; single implementation; RC-C evidence-tier sensitivity	OUT	RC-C	Terminal human-agency exemplar
A017	No independent validation identified	HIGH	HIGH	Exact technical artifact; action-class and governance controls	Technical self-report; no narrow human-control outcome	OUT	RC-A	Governance mechanism evidence

Formal-publication-channel sensitivity

Restriction	n	RC-A	RC-B	RC-C	Outcome-evaluable	Local control > global stop	Attention recruitment clear	Direct human exercise	Interpretation
Formal journal/conference publication channel only	8	5	3	0	7	8/8	6/8	1/7 (base and strict HIGH)	Core local/selective-control and system-mediated-attention directions persist. RC-C disappears, and direct-human-exercise evidence becomes substantially thinner. This is a credibility sensitivity, not a superior-evidence estimator.

Appraisal principle: no composite quality score is calculated. Publication form, directness, deployment evidence, independence, self-report dependence, source traceability, and outcome attribution are intentionally kept separate. 'Formal publication channel' denotes journal/conference publication status; it is not treated as proof of methodological validity, nor is technical evidence treated as invalid solely because it was not journal-published.

Supplementary Table S6

Competing-concept differentiation map for Control Assimilation

Panel A. Prior-art contribution, oversight timing, authority, and human cognition

Reference / framework	Primary analytic unit	Core contribution already established in prior art	Routing / timing of oversight	Formal authority vs effective control	Human detection / cognition
Bainbridge 1983 — Ironies of automation	Human operator + automated process	Automation can remove humans from routine control while leaving difficult abnormal-event responsibility.	Indirect: takeover occurs after automation has shaped operator engagement.	Implicit responsibility/control mismatch.	Out-of-the-loop skill/awareness problem anticipated.
Endsley & Kiris 1995 — OOTL performance	Operator performance under automation levels	Automation level can reduce situation awareness and degrade takeover performance.	Takeover timing is central but not modeled as a cross-system routing architecture.	Not the primary construct.	Directly measures situation awareness / takeover performance.
Santoni de Sio & van den Hoven 2018 — Meaningful Human Control	Normative human-system responsibility relation	Meaningful control requires tracing responsibility and tracking relevant human reasons.	Temporal/proximity implications but not an operational routing metrology.	Strong normative treatment.	Not primarily a human-detection model.
Sterz et al. 2024 — effective human oversight	Interdisciplinary oversight conditions	Effective oversight requires causal power and suitable epistemic access, among other conditions.	Recognizes access and intervention conditions.	Explicitly separates nominal oversight from effective oversight.	Human capabilities and epistemic access are relevant.
Langer et al. 2025 — signal-detection perspective	Human overseer detecting AI errors	Reliable error detection is crucial for effective human oversight; SDT separates sensitivity and response bias.	Timing is secondary to detection effectiveness.	Oversight effectiveness, not formal-authority architecture.	Central analytic focus.
Tsamados, Floridi & Taddeo 2025 — supervision to teaming	Operational human control / collaborative agency	For AI systems based on foundation models, human control is better understood through human-machine teaming and collaborative agency than through exclusive human supervision.	Discusses ongoing control and human interventions.	Control is distributed across interacting agents.	Human capabilities and system interplay discussed.
Nguyen et al. 2026 — controllability in agentic AI	Controllability paradigms / survey literature	Constraints/guardrails, adaptive control, Agent-in-the-Loop and Human-in-the-Loop paradigms; supervisory agents and escalation are prior art.	Explicit discussion of escalation, supervisor independence and access.	Distinguishes causal and normative control.	Not a human-detection experiment.
Zhu et al. 2026 — meaningful human oversight design	End-to-end oversight design patterns	Separates AI operative agency from human evaluative agency and maps design/runtime oversight mechanisms.	Explicitly relevant to runtime oversight design.	Strong distinction between operative/evaluative roles.	Human evaluative agency emphasized.
Takahashi 2026 — oversight-centered metrology	Oversight channels in deployed agentic workflows	Humans and automated verifiers are costly interrupt channels characterized by error, delay, availability, congestion, routing error, selective review and irreversibility.	Central analytic focus; especially close prior art.	Human authority is not the primary unit, but intervention channels determine real control.	Models human review limits without privileged-oracle assumption.

Reference / framework	Primary analytic unit	Core contribution already established in prior art	Routing / timing of oversight	Formal authority vs effective control	Human detection / cognition
Estep 2026 — instrumental succession	Longitudinal theory of agency transfer	Human controllers may progressively pursue instrumental dispositions that increase AI's share of oversight/control, potentially shifting the locus of agency.	Not primarily an oversight-routing theory.	Agency/control locus is central.	Not primarily cognitive-performance measurement.
Green 2022 — flaws of mandatory human oversight	Policy/organizational human review	Nominal human oversight can provide false assurance when humans cannot perform the expected corrective role.	Recognizes practical constraints on review.	Central critique of nominal vs effective oversight.	Human limitations relevant.
VDE VCIO 2022 — Responsible Human Oversight assurance	Socio-technical AI trustworthiness assessment	Operationalizes HIC/HITL/HOTL-related assurance through operator expertise, interaction effort, documented human control/involvement, human takeover, and post-hoc transparency/change.	Addresses monitoring and takeover readiness, but does not use upstream selection/routing of control-relevant states as the primary analytic unit.	Explicitly evaluates human control/involvement and takeover conditions.	Operator expertise, intelligibility, and interaction effort are explicit indicators.

Panel B. Agentic runtime focus, empirical reconstruction, overlap, residual distinction, and novelty consequence

Reference / framework	Agentic or autonomous runtime focus	Cross-system empirical architecture reconstruction	Primary overlap with Control Assimilation	Residual distinction retained by this study	Novelty consequence	Manuscript reference
Bainbridge 1983 — Ironies of automation	Classical automation; not agentic AI.	No	Foundational antecedent for downstream human supervisory position.	CA reconstructs the technical path that determines when/how the responsible human reaches a control point across concrete agentic architectures.	No novelty claim for the general irony of automation or OOTL problem.	Bainbridge, 1983
Endsley & Kiris 1995 — OOTL performance	Automation, not contemporary agentic architectures.	No	Shows that retained takeover authority may not imply effective control.	CA separates opportunity, capability, authority, and observed exercise while reconstructing upstream mediation.	No novelty claim for takeover-performance degradation.	Endsley & Kiris, 1995
Santoni de Sio & van den Hoven 2018 — Meaningful Human Control	Autonomous systems broadly.	No	Formal authority alone is insufficient for meaningful control.	CA is descriptive/empirical rather than normative: it maps actual runtime control paths and upstream technical mediation.	MHC is prior art; CA must not be presented as a replacement normative theory.	Santoni de Sio & van den Hoven, 2018
Sterz et al. 2024 — effective human oversight	AI systems broadly.	No	Strong prior art for architecture/outcome separation.	CA operationalizes a cross-architecture path reconstruction and keeps observed human exercise separate from governance architecture.	No novelty claim for causal power / epistemic access as requirements.	Sterz et al., 2024

Reference / framework	Agentic or autonomous runtime focus	Cross-system empirical architecture reconstruction	Primary overlap with Control Assimilation	Residual distinction retained by this study	Novelty consequence	Manuscript reference
Langer et al. 2025 — signal-detection perspective	AI-based systems at runtime/inspection time.	No	Shows that a human monitoring role is insufficient if relevant errors are not reliably detected.	CA examines the upstream technical path that determines which cases/information reach the human before detection can occur.	No novelty claim for human error-detection effectiveness or SDT framing.	Langer et al., 2025
Tsamados, Floridi & Taddeo 2025 — supervision to teaming	Foundation-model-based AI; defence/security examples.	Conceptual/review, not CA-style path reconstruction.	Direct competitor to simple supervisory-control framing.	CA does not claim exclusive human supervision; it empirically reconstructs where the responsible-human boundary sits and what is upstream of it.	No novelty claim that collaborative/distributed control is preferable to exclusive supervision for foundation-model-based AI.	Tsamados et al., 2025
Nguyen et al. 2026 — controllability in agentic AI	Directly agentic AI.	Structured survey, not empirical reconstruction of concrete heterogeneous runtime paths.	Strong competitor for control taxonomy, AI-on-AI oversight, escalation and guardrails.	CA's residual contribution is provenance-preserving empirical architecture reconstruction using the responsible-human control path as unit, plus observed human-exercise attribution.	No novelty claim for AITL/HITL, supervisory agents, guardrails, or escalation paths.	Nguyen et al., 2026
Zhu et al. 2026 — meaningful human oversight design	Contemporary AI systems.	Design framework rather than CA-style empirical cross-system path synthesis.	Strong competitor for function decomposition and end-to-end oversight framing.	CA provides case-level path reconstruction, MMCD/path coding, RC boundary configurations, and outcome-attribution categories.	No novelty claim for operative/evaluative agency distinction or layered oversight design.	Zhu et al., 2026
Takahashi 2026 — oversight-centered metrology	Directly agentic/deployment-relevant.	Theoretical/metrological framework rather than provenance-preserving empirical reconstruction of heterogeneous concrete architectures.	Closest competitor for the claim that oversight access, routing and latency are architecturally conditioned.	CA cannot claim novelty for routing/latency/scarcity review. Its distinct contribution is cross-architecture empirical reconstruction of formal authority plus upstream technical mediation, RC boundary location, ICC, and observed human exercise.	Materially narrows novelty; strengthens the residual empirical architecture claim.	Takahashi, 2026
Estep 2026 — instrumental succession	Advanced AI / societal agency.	No	Adjacent trajectory-level account of human-to-AI control transfer.	CA is structural and cross-sectional: it can exist while formal authority remains human and does not require observed temporal succession of agency.	Prevents interpreting 'assimilation' as an empirically demonstrated longitudinal progression.	Estep, 2026

Reference / framework	Agentic or autonomous runtime focus	Cross-system empirical architecture reconstruction	Primary overlap with Control Assimilation	Residual distinction retained by this study	Novelty consequence	Manuscript reference
Green 2022 — flaws of mandatory human oversight	Government algorithms, broader than agentic AI.	No	Direct antecedent to formal-authority/effective-control divergence.	CA identifies and codes a specific runtime architecture mechanism by which practical access can be technically conditioned.	No novelty claim that nominal oversight can fail.	Green, 2022
VDE VCIO 2022 — Responsible Human Oversight assurance	AI systems/applications broadly; not specific to agentic AI.	No	Practical assurance comparator for whether nominal human oversight is operationally supported.	CA adds a path-level question: whether the upstream architecture that detects, filters, classifies, represents, or routes control-relevant states to the human is itself examined.	No novelty claim for HIC/HITL/HOTL assurance, takeover readiness, operator expertise, or documented human involvement.	VDE, 2022

Interpretive boundary: Supplementary Table S6 is a differentiation map, not a novelty score. Several components of Control Assimilation are explicit prior art. The residual contribution claimed in the manuscript is the provenance-preserving empirical reconstruction of the responsible human's runtime control path across heterogeneous real/deployment-proximal architectures, with formal authority, intervention opportunity/capability, machine-mediated path structure, independent constraints, and observed human-control exercise kept analytically separate. The 17-case cohort is a frozen first-wave analytic sample, not a probability sample; counts and configuration proportions are descriptive and not population-prevalence estimates.

Reference cited in Supplementary Table S6

Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., & Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. arXiv (Version 2). <https://doi.org/10.48550/arXiv.2404.04059>. Related ACM FAccT 2024 publication: <https://doi.org/10.1145/3630106.3659051>.