

JULIEN MARSCHALL  
DOCTORAL DISSERTATION

# HUMAN × AI × REALITY

EIN OPERATOR IM EXPERIMENT

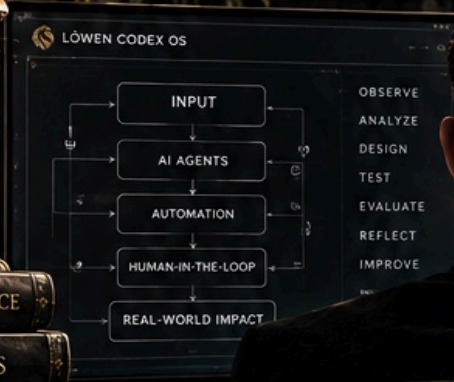
EINE LONGITUDINALE, INTERDISZIPLINÄRE UND  
DESIGN-SCIENCE-RESEARCH-ANALYSE ZUR WIRKSAMKEIT,  
GOVERNANCE UND SKALIERBARKEIT EINES AI-NATIVEN  
OPERATING SYSTEMS IN DER REALWIRTSCHAFT

SYSTEME  
MENSCHEN  
GESELLSCHAFT  
WIRKUNG  
ZUKUNFT

VON FRAGEN  
ZU BEWEISBAREN  
ANTWORTEN.

„NICHT MEHR  
MÖGLICHKEITEN.  
SONDERN BESSERE  
SYSTEME.“

*Julien Marschall*



ARTIFICIAL INTELLIGENCE

HUMAN FACTORS

SYSTEMS ENGINEERING

ENTREPRENEURSHIP

SOCIAL IMPACT



SAME MIND  
HIGHER STANDARDS

## LÖWEN CODEX OS

WISSENSCHAFT. PRAXIS. WIRKUNG.

VERSION 1.0  
DOCTORAL DISSERTATION  
RESEARCH EDITION  
SEPTEMBER 2026

DESIGN SCIENCE RESEARCH  
×  
REAL-WORLD VALIDATION

CODEx RESEARCH SERIES  
LC-DR-JM-001-V01-2026  
LÖWEN CODEX  
INDEPENDENT RESEARCH  
REAL SYSTEMS. REAL EVIDENCE.

# PUBLIC RESEARCH EDITION

## HUMAN x AI x REALITY - ARBEIT 2

Diese Ausgabe erhaelt den wissenschaftlichen Haupttext, Forschungsfragen, Theorie, Methodik, Evaluation, Negativergebnisse, Limitationen und die abstrahierten Designprinzipien der Arbeit. Reproduktionskritische technische Einzelheiten wurden gezielt entfernt oder abstrahiert. Dazu gehoeren insbesondere konkrete Zugangspfade, interne Endpunkte, Ports, Datei- und Datenbankstrukturen, Hash-/Registerdetails, Provider- und Accountkonfigurationen sowie Implementierungsdetails, deren Kombination eine Rekonstruktion des produktiven Loewen Codex OS erleichtern wuerde.

Die Reduktion ist eine Disclosure-Grenze, keine nachtraegliche Aenderung der empirischen Befunde. Wo ein Detail fuer die Nachvollziehbarkeit des Forschungsclaims erforderlich ist, bleibt die Aussage auf funktionaler oder methodischer Ebene erhalten. Die vollstaendige technische Evidenz verbleibt in der nicht oeffentlichen Research Edition.

---

Disclosure-Prinzip

PUBLIC: Forschungsdesign, theoretischer Rahmen, Evidenzlogik, Ergebnisstruktur, Grenzen und Designwissen.

PRIVATE: reproduktionskritische Systemarchitektur, konkrete Betriebswege, interne Konfiguration, technische Angriffsflaechen und operative Detaildaten.

# HUMAN x AI x REALITY

Longitudinale Evaluation eines AI-nativen Operator-Systems · ARBEIT 2 · Private  
Deep Edition

Julien Marschall

Version 1.0 · Final Research Edition · 20 September 2026 · Independent  
Longitudinal Research Monograph · nicht peer reviewed / external review  
pending · Study Freeze 2026-09-20T08:46:41

# Inhaltsverzeichnis

|                                                                                                                    |           |
|--------------------------------------------------------------------------------------------------------------------|-----------|
| <b>ARBEIT 2 · PRIVATE DEEP FINAL · 2026-09-20</b>                                                                  | <b>5</b>  |
| <b>FRONT MATTER MASTER v1 · ARBEIT 2 · 2026-09-20</b>                                                              | <b>6</b>  |
| Deckblatt . . . . .                                                                                                | 6         |
| Research Status . . . . .                                                                                          | 6         |
| Transparency Statement . . . . .                                                                                   | 6         |
| Authorship / Role Statement . . . . .                                                                              | 7         |
| Generative AI Statement . . . . .                                                                                  | 7         |
| Study Relationship Statement: Work 1 -> Work 2 . . . . .                                                           | 7         |
| Abstract (DE) . . . . .                                                                                            | 7         |
| Abstract (EN) . . . . .                                                                                            | 8         |
| Research Question Summary . . . . .                                                                                | 8         |
| Freeze / Version Statement . . . . .                                                                               | 8         |
| Data / Evidence Statement . . . . .                                                                                | 9         |
| Table of Contents (Quelle: S4 Manuskript Core, S5 Appendix Master; keine Seitenzahlen<br>vor dem Layout) . . . . . | 9         |
| Appendix Overview . . . . .                                                                                        | 9         |
| <b>HUMAN x AI x REALITY - Longitudinale Folgeuntersuchung des Loewen Codex OS</b>                                  | <b>11</b> |
| ARBEIT 2 · MANUSKRIFT CORE v1 · Kapitel I-XIV · 2026-09-20 . . . . .                                               | 11        |
| <b>KAPITEL I · INTRODUCTION</b>                                                                                    | <b>12</b> |
| 1.1 Problemstellung . . . . .                                                                                      | 12        |
| 1.2 Forschungsbedarf . . . . .                                                                                     | 12        |
| 1.3 Work 1 -> Work 2 Motivation . . . . .                                                                          | 12        |
| 1.4 Forschungsfragen . . . . .                                                                                     | 12        |
| 1.5 Forschungsziel . . . . .                                                                                       | 13        |
| 1.6 Beitrag . . . . .                                                                                              | 13        |
| 1.7 Scope . . . . .                                                                                                | 13        |
| 1.8 Aufbau der Arbeit . . . . .                                                                                    | 13        |
| <b>KAPITEL II · THEORY AND RELATED WORK</b>                                                                        | <b>14</b> |
| 2.1 Design Science Research . . . . .                                                                              | 14        |
| 2.2 Action Design Research als ergaenzende Perspektive . . . . .                                                   | 14        |
| 2.3 Human-AI Interaction . . . . .                                                                                 | 14        |
| 2.4 Automation, Human Oversight und Trust . . . . .                                                                | 14        |
| 2.5 AI Governance . . . . .                                                                                        | 15        |
| 2.6 Reliability Engineering . . . . .                                                                              | 15        |
| 2.7 Fail-safe, Fail-closed, Default-deny . . . . .                                                                 | 15        |
| 2.8 Operational Observability . . . . .                                                                            | 15        |
| 2.9 Provenance, Auditability, Reproducibility . . . . .                                                            | 16        |
| 2.10 Measurement Reliability und Instrumentation Threats . . . . .                                                 | 16        |
| 2.11 Agent Architectures (Abgrenzung) . . . . .                                                                    | 16        |
| 2.12 Analytisches Rahmenmodell . . . . .                                                                           | 16        |
| <b>KAPITEL III · RESEARCH DESIGN AND METHODOLOGY</b>                                                               | <b>17</b> |
| 3.1 DSR Study Design . . . . .                                                                                     | 17        |
| 3.2 Longitudinal N-of-1 Design . . . . .                                                                           | 17        |
| 3.3 Artefact / Operator Units . . . . .                                                                            | 17        |

|                                                                                  |           |
|----------------------------------------------------------------------------------|-----------|
| 3.4 Phasen . . . . .                                                             | 17        |
| 3.5 Data Sources . . . . .                                                       | 17        |
| 3.6 Evidence Classes . . . . .                                                   | 18        |
| 3.7 Maturity Model . . . . .                                                     | 18        |
| 3.8 HITL-P / HITL-R / HITL-B . . . . .                                           | 18        |
| 3.9 Evaluation Methods . . . . .                                                 | 18        |
| 3.10 Claim Analysis Procedure . . . . .                                          | 18        |
| 3.11 Literature Verification . . . . .                                           | 18        |
| 3.12 Validity . . . . .                                                          | 18        |
| 3.13 Ethics / Privacy / Legal Boundary . . . . .                                 | 18        |
| 3.14 Reproducibility . . . . .                                                   | 19        |
| <b>KAPITEL IV • FOUNDATIONAL STUDY AND TRANSITION FROM WORK 1</b>                | <b>20</b> |
| <b>KAPITEL V • AS-BUILT ARTEFACT ARCHITECTURE AT WORK-2 FREEZE</b>               | <b>27</b> |
| 5.1 Architekturebenen . . . . .                                                  | 27        |
| 5.2 Komponenten und Maturity . . . . .                                           | 27        |
| 5.3 State Sources und Entities . . . . .                                         | 27        |
| 5.4 Caller Workbench, Contact Points, Manual Mail . . . . .                      | 27        |
| 5.5 manual_send_requests (F3-Erratum) . . . . .                                  | 28        |
| 5.6 Harness und Runtime . . . . .                                                | 28        |
| 5.7 Public Representation . . . . .                                              | 28        |
| <b>KAPITEL VI • STATE, GOVERNANCE, HUMAN AUTHORITY AND EVIDENCE ARCHITECTURE</b> | <b>29</b> |
| 6.1 Authoritative vs Reported State . . . . .                                    | 29        |
| 6.2 Governance Gates . . . . .                                                   | 29        |
| 6.3 Human Authority vs Agent Execution . . . . .                                 | 29        |
| 6.4 Verification Model A-E . . . . .                                             | 29        |
| 6.5 Evidence Architecture und Claim Scope . . . . .                              | 29        |
| 6.6 Nicht-Gleichsetzungen . . . . .                                              | 30        |
| <b>KAPITEL VII • LONGITUDINAL PHASE MODEL</b>                                    | <b>31</b> |
| 7.1 Historischer Kontext und T0 . . . . .                                        | 31        |
| 7.2 Pre-T1 (23.08.-06.09.) . . . . .                                             | 31        |
| 7.3 T1 (ab 2026-09-06T22:56:02) . . . . .                                        | 31        |
| 7.4 Extension (14.-18.09.) . . . . .                                             | 31        |
| 7.5 Post-T1-Build (18.-20.09.) . . . . .                                         | 31        |
| 7.6 CURRENT STUDY FREEZE . . . . .                                               | 31        |
| <b>KAPITEL VIII • EVALUATION BY EVIDENCE TYPE</b>                                | <b>33</b> |
| Static / Structural Checks . . . . .                                             | 33        |
| Unit / Integration Tests . . . . .                                               | 33        |
| Counterfactual Tests . . . . .                                                   | 33        |
| Regression Tests . . . . .                                                       | 33        |
| Runtime Verification . . . . .                                                   | 33        |
| Delivered Verification . . . . .                                                 | 34        |
| Manifest / Chain . . . . .                                                       | 34        |
| Scoped Browser Acceptance . . . . .                                              | 34        |
| Zero-Write . . . . .                                                             | 34        |
| Zusammenfassung und Measurement Change . . . . .                                 | 34        |

|                                                                                                             |           |
|-------------------------------------------------------------------------------------------------------------|-----------|
| <b>KAPITEL X · HUMAN-AI, HITL, INTERVENTIONS, REGISTRATION</b>                                              | <b>37</b> |
| 10.1 HITL-Register . . . . .                                                                                | 37        |
| 10.2 Label ist nicht Handlung . . . . .                                                                     | 37        |
| 10.3 Acceptance als menschliche Abnahme . . . . .                                                           | 37        |
| 10.4 Interventionen und Registrierung . . . . .                                                             | 37        |
| <b>KAPITEL XI · OPERATIONAL AND BUSINESS OUTCOMES</b>                                                       | <b>38</b> |
| <b>KAPITEL XII · DISCUSSION</b>                                                                             | <b>39</b> |
| RQ0 - Beschreibbarkeit und Artefaktgrenze (ANSWERED DESCRIPTIVELY) . . . . .                                | 39        |
| RQ1 - Observability / State Verification (ANSWERED DESCRIPTIVELY) . . . . .                                 | 39        |
| RQ2 - Governance / Enforcement (PARTIALLY ANSWERED) . . . . .                                               | 40        |
| RQ3 - Human-AI / HITL (PARTIALLY ANSWERED) . . . . .                                                        | 40        |
| RQ4 - Longitudinal Change (PARTIALLY ANSWERED) . . . . .                                                    | 40        |
| RQ5 - Effectiveness (NOT MEASURED) . . . . .                                                                | 41        |
| RQ6 - Representation Fidelity (ANSWERED DESCRIPTIVELY) . . . . .                                            | 41        |
| Sieben Ungleichungen . . . . .                                                                              | 41        |
| Instrumentfehler und Measurement Change . . . . .                                                           | 42        |
| Fail-closed als Design Transfer . . . . .                                                                   | 42        |
| Neue Widersprueche . . . . .                                                                                | 42        |
| <b>KAPITEL XIII · CANDIDATE DESIGN PRINCIPLES</b>                                                           | <b>43</b> |
| DP-C1 · AUTHORITATIVE STATE SHOULD REMAIN SERVER-SIDE FOR CONSEQUENTI-<br>AL ACTIONS . . . . .              | 43        |
| DP-C2 · PREVIEW AND EXECUTION SHOULD BE SEPARATE STATES . . . . .                                           | 43        |
| DP-C3 · IDENTITY SHOULD BE BOUND MORE STRONGLY THAN HEURISTIC MATCHES<br>BEFORE EXTERNAL ACTION . . . . .   | 43        |
| DP-C4 · EVIDENCE SCOPE SHOULD TRAVEL WITH THE CLAIM . . . . .                                               | 44        |
| DP-C5 · VERIFICATION SHOULD USE MULTIPLE INDEPENDENT EVIDENCE LAYERS . . . . .                              | 44        |
| DP-C6 · HUMAN AUTHORITY SHOULD BE DISTINCT FROM AGENT EXECUTION . . . . .                                   | 44        |
| DP-C7 · MEASUREMENT INSTRUMENTS SHOULD BE VERSIONED AND TREATED AS<br>PART OF THE RESEARCH SYSTEM . . . . . | 45        |
| DP-C8 · FREEZES SHOULD SEPARATE EVALUATION FROM CONTINUOUS DEVELOP-<br>MENT . . . . .                       | 45        |
| <b>KAPITEL XIV · LIMITATIONS, CONCLUSION, FUTURE RESEARCH</b>                                               | <b>46</b> |
| 14.1 Limitations . . . . .                                                                                  | 46        |
| 14.2 Final RQ Answers . . . . .                                                                             | 47        |
| 14.3 Contributions . . . . .                                                                                | 48        |
| 14.4 Conclusion . . . . .                                                                                   | 48        |
| 14.5 Future Research . . . . .                                                                              | 50        |
| 14.6 T2 Boundary . . . . .                                                                                  | 50        |
| <b>LITERATURVERZEICHNIS</b>                                                                                 | <b>51</b> |
| <b>ARBEIT 2 · APPENDICES MASTER v1 · 2026-09-20</b>                                                         | <b>54</b> |
| Master Index . . . . .                                                                                      | 54        |
| APPENDIX A · Foundational Study / Work 1 . . . . .                                                          | 57        |
| APPENDIX B · Claim Matrix v2 . . . . .                                                                      | 59        |
| APPENDIX C · Methodology / Timeline . . . . .                                                               | 59        |
| APPENDIX D · As-built Artefact Inventory . . . . .                                                          | 59        |
| APPENDIX E · Evaluation Method Matrix . . . . .                                                             | 60        |
| APPENDIX F · Failure Cases . . . . .                                                                        | 60        |
| APPENDIX G · RQ Evidence / Traceability (Master-Tabelle) . . . . .                                          | 60        |
| APPENDIX H · HITL / Human Decision Evidence . . . . .                                                       | 62        |
| APPENDIX I · Interventions / Incidents (Summary; Register maschinenlesbar, PRIVA-<br>TE ONLY) . . . . .     | 62        |
| APPENDIX J · Deploy / Manifest / Acceptance / Chain Evidence . . . . .                                      | 63        |
| APPENDIX K · Negative / Null / Not-Estimable Results . . . . .                                              | 63        |
| APPENDIX L · Instrument Registry . . . . .                                                                  | 67        |
| APPENDIX M · Phase State Table . . . . .                                                                    | 67        |
| APPENDIX N · Literature / Theory Alignment . . . . .                                                        | 67        |

|                                                              |    |
|--------------------------------------------------------------|----|
| APPENDIX O · Current Study Freeze . . . . .                  | 69 |
| APPENDIX P · Reproducibility / Publication Package . . . . . | 70 |
| Crosschecks . . . . .                                        | 70 |

# **ARBEIT 2 · PRIVATE DEEP FINAL · 2026-09-20**

# FRONT MATTER MASTER v1 · ARBEIT 2 · 2026-09-20

Neutraler Master (kein Public/Private-Split; S8). Generator S6-1.0 (2026-09-20). Kein akademischer Grad, kein Peer Review, keine DOI, keine Lizenz - noch nicht vergeben.

## Deckblatt

### **HUMAN x AI x REALITY**

Longitudinale Evaluation eines AI-nativen Operator-Systems. Eine Design-Science-Research-Studie zu Governance, Human Authority, Zustandsprovenienz, Failure Semantics und evidenzbasierter Verifikation

*Longitudinal Evaluation of an AI-native Operator System. A Design Science Research Study on Governance, Human Authority, State Provenance, Failure Semantics and Evidence-based Verification*  
Version 1.0 · Final Research Edition · 20 September 2026

Autor: Julien Marschall

Institutionelle Zugehoerigkeit: Unabhaengiger Forscher / Entwickler (Independent Researcher / Developer)

Status: Independent Longitudinal Research Monograph · Review Status: nicht peer reviewed / external review pending

MANUSCRIPT\_VERSION: Version 1.0 · Final Research Edition · 20 September 2026 (interne Provenienz: v1 -> v2 PATCHED -> S8 -> S9 FINAL CANDIDATE -> FINAL)

GENERATOR\_VERSION: je Artefakt (F1-1.0 ... S7P-1.0, Appendix L)

CURRENT\_STUDY\_FREEZE / DATA BOUNDARY: 2026-09-20T08:46:41 (Record-Hash 5c1c6244...)

ARTEFACT\_VERSION: Build DEPLOY-20260920-0719-LOG0506 (a98e51b632cc)

Study Freeze: 20. September 2026, 08:46:41 (CURRENT STUDY FREEZE)

Artefakt: Loewen Codex OS

Vorgaengerstudie: Work 1 - Independent Research Monograph, Artefaktfreeze 2026-08-22

## Research Status

Diese Arbeit ist eine Unabhaengige longitudinale Forschungsmonographie (Independent Longitudinal Research Monograph), Version 1.0 · Final Research Edition · 20 September 2026, Review Status: nicht peer reviewed / external review pending (kein dokumentierter externer Review zum Zeitpunkt der Finalisierung), optional bezeichnet als Longitudinal Follow-up Study of the Loewen Codex OS. Sie entstand ausserhalb eines institutionellen Promotions- oder Studiengangs, ist keine universitaere Dissertation, und mit ihr ist kein akademischer Grad verbunden. Die Bezeichnung Research Monograph beschreibt den wissenschaftlich strukturierten Forschungsoutput. Eine externe Begutachtung ist moeglich und geplant; nur dokumentierte Reviews duerfen spaeter genannt werden. Interne Projektbezeichnungen (Niveau-Zuschreibungen) gehoeren nicht auf das Deckblatt.

## Transparency Statement

Autor, Entwickler, primaerer Operator, Forschender, teilweise Tester und Acceptance Authority sind dieselbe Person. Daraus entstehen Bias-Risiken (selektive Wahrnehmung, nachtraegliche Ra-

tionalisierung, Selektion durch Testfokus). Methodische Gegenmassnahmen: Freeze mit Record-Hash, append-only Evidenz mit Hash-Ketten, Claim-Matrix mit Status je Aussage, maschinell erzeugte Tabellen aus Primaerdaten, Counterfactual-Tests mit Gegenprobe, explizit gefuehrte Counterevidence, scope-begrenzte Abnahme, Literaturverifikation gegen Originalquellen. Diese Massnahmen reduzieren das Risiko; sie eliminieren den Bias nicht.

## **Authorship / Role Statement**

Alleiniger Autor: Julien Marschall (Konzeption, Systembau, Betrieb, Datenerhebung, Analyse, Manuskript, finale Verantwortung). Eine zweite Operatorrolle (Caller) ist im Methodik- und Artefakttext neutral beschrieben; sie ist system user / operator role / data source, nicht co-author, nicht co-researcher und nicht als eigenstaendig instrumentierte Studienteilnehmerin gefuehrt. Externe Gutachter-Ruecklaeufer (KI-gestuetzt, E1/REPORTED) wurden als Konsistenz-Input behandelt, nicht als Primaerevidenz.

## **Generative AI Statement**

Generative KI (Claude-Modelle in Cowork/Claude Code; ChatGPT fuer Gutachter-Ruecklaeufer) wurde eingesetzt fuer Strukturierung, Textentwurf, Redaktion, den Bau der Code- und Artefaktgeneratoren (F-1..F-6, S1, S3-S6), Literaturorientierung und -verifikationsunterstuetzung sowie Konsistenzpruefung. Sie war nicht empirische Primaerquelle: Systemzustaende, Zahlen und Ergebnisse stammen ausschliesslich aus registrierten Evidenzquellen (Register, Manifeste, Logs, Zustandsartefakte, Test-/Acceptance-Evidenz, Freeze-Records). Es wird nicht behauptet, dass KI die Arbeit unabhaengig wissenschaftlich validiert hat. Die finale Verantwortung liegt beim Autor.

## **Study Relationship Statement: Work 1 -> Work 2**

WORK 1 (Freeze 2026-08-22): retrospektive Foundational Study - Artefakt- und Prozessrekonstruktion, zehn Designprinzipien als Hypothesen. WORK 2 (Freeze 2026-09-20T08:46:41): longitudinale Follow-up Study mit prospektiven Elementen, Failure Semantics, Measurement Reliability, Human Decision Provenance, Scoped Acceptance und Freeze Boundary. Work 1 dient als historischer Ausgangspunkt; es ist keine Primaerevidenz fuer Septemberzustaende. Die empirischen Korpora bleiben getrennt (Augustbefund bleibt Augustbefund). Crosswalk: Appendix A.

## **Abstract (DE)**

Lebende Human-AI-Systeme, die ein einzelner Betreiber entwickelt, betreibt und zugleich erforscht, sind schwer belastbar zu beschreiben: Anzeigen sind nicht ihr Zustand, Kontrollen gelten nur dort, wo sie eingebaut wurden, und die Messinstrumente sind Teil des untersuchten Systems. Diese unabhaengige longitudinale Forschungsmonographie untersucht das Loewen Codex OS - ein AI-natives Operator-System mit lokalen Sprachmodellen, deterministischer Automatisierung, Telefonie, Content-Publishing und einem Caller-Arbeitsplatz - als Folgestudie zur retrospektiven Foundational Study (Work 1, Freeze 2026-08-22). Das Design ist eine longitudinale Design-Science-Fallstudie (N-of-1, ein Principal Operator, keine Kontrollbedingung) mit den Phasen T0, Pre-T1, T1 (ab 2026-09-06T22:56:02), praeregistrierter Extension, Post-T1-Build und CURRENT STUDY FREEZE (2026-09-20T08:46:41). Evaluiert wurde nach Evidenzarten: statische Pruefungen, Unit-/Integrationstests, Counterfactuals, Regression, Runtime- und Delivered-Verifikation, Deploy-Manifeste mit Hash-Kette, scope-begrenzte Browser-Abnahme und Zero-Write-Messung; ergaenzt um Failure Cases, HITL-Register und Interventions-/Incident-Register. Jede Aussage traegt Population, Zeitfenster, Evidenzklasse und Status. Ergebnis: Mehrere Governance- und Verifikationsmechanismen - fail-closed Freigaben fuer extern wirksame Aktionen, serverseitig gebundene Identitaet, Preview-Invalidation, Sender-Bindung - waren in definierten Pfaden funktional testbar; 32 Gegenproben diskriminierten in den geprueften Faellen zwischen altem und neuem Verhalten. Zugleich dokumentiert die Studie 16 Failure Cases, darunter Instrumentierungsfehler des eigenen Harness, Governance-Inhomogenitaet zwischen Nachbarpfaden, unvollstaendige HITL-Abdeckung (579 klassifizierte von 3.224 Registerzeilen; registrierte Labels sind keine bestaetigten Reviews) und Outcome-Dimensionen, die mit diesem Design nicht messbar waren:

operative und oekonomische Wirksamkeit wurden nicht gemessen, Realversand wurde nicht getestet, die rechtliche Bewertung bleibt offen. Beitrag ist eine fallbezogene Evidenzarchitektur, die Artefaktzustand, Messzustand, Evidenzzustand und Claim-Scope trennt, Failure Cases einschliesslich Messfehlern fuehrt und acht Candidate Design Principles zur kuenftigen Validierung vorschlaegt. Die Befunde gelten fuer das untersuchte System im eingefrorenen Zeitraum; sie belegen keine generalisierbare, operative, oekonomische oder rechtliche Wirksamkeit.

Schluesselwoerter: Design Science Research; Mensch-KI-Interaktion; KI-Governance; Human-in-the-Loop; Zuverlaessigkeitstechnik; Operational Observability; Zustandsprovenienz; AI-native Informationssysteme; Fail-Closed-Governance; Longitudinale Fallstudie

## Abstract (EN)

Living human-AI systems that a single operator develops, runs and studies at the same time are hard to describe reliably: what they display is not their state, their controls apply only where they were built in, and the measurement instruments are part of the system under study. This independent longitudinal research monograph examines the Loewen Codex OS - an AI-native operator system combining local language models, deterministic automation, telephony, content publishing and a caller workbench - as a follow-up to the retrospective foundational study (Work 1, freeze 2026-08-22). The design is a longitudinal design science case study (N-of-1, one principal operator, no control condition) with the phases T0, pre-T1, T1 (from 2026-09-06T22:56:02), a pre-registered extension, a post-T1 build and the CURRENT STUDY FREEZE (2026-09-20T08:46:41). Evaluation proceeded by evidence type: static checks, unit/integration tests, counterfactuals, regression, runtime and delivered verification, deploy manifests with a hash chain, scoped browser acceptance and zero-write measurement; complemented by failure cases, a HITL register and intervention/incident registers. Every statement carries population, time window, evidence class and status. Results: several governance and verification mechanisms - fail-closed release of externally consequential actions, server-bound identity, preview invalidation, sender binding - were functionally testable in defined paths; 32 counterfactual checks discriminated between old and new behaviour in the cases examined. At the same time the study documents 16 failure cases, including instrumentation faults of its own harness, governance inhomogeneity between neighbouring paths, incomplete HITL coverage (579 classified of 3,224 register rows; registered labels are not confirmed reviews) and outcome dimensions that could not be established under the present design: operational and economic effectiveness were not measured, real sending was not tested, and the legal assessment remains open. The contribution is a case-derived evidence architecture that separates artefact state, measurement state, evidence state and claim scope, records failure cases including measurement faults, and proposes eight candidate design principles for future validation. The findings hold for the system studied within the frozen period; they do not establish generalisable, operational, economic or legal effectiveness.

Keywords: Design Science Research; Human-AI Interaction; AI Governance; Human-in-the-Loop; Reliability Engineering; Operational Observability; State Provenance; AI-native Information Systems; Fail-Closed Governance; Longitudinal Case Study

## Research Question Summary

- RQ0: ANSWERED DESCRIPTIVELY
- RQ1: ANSWERED DESCRIPTIVELY
- RQ2: PARTIALLY ANSWERED
- RQ3: PARTIALLY ANSWERED
- RQ4: PARTIALLY ANSWERED
- RQ5: NOT MEASURED / NOT ANSWERABLE WITH CURRENT DESIGN
- RQ6: ANSWERED DESCRIPTIVELY  
(keine Quote 'n von 7 beantwortet')

## Freeze / Version Statement

Studie, T2 (DRAFT) oder einen klar getrennten Post-Freeze-Kontext. Manuskriptversion Version 1.0 · Final Research Edition · 20 September 2026 (FINAL = interner Forschungs- und Manuskriptabschluss dieser Version; nicht peer reviewed, nicht publiziert, kein akademischer Grad); Generator-/Artefaktversionen sind in den Metadaten und der Instrument Registry (Appendix L) verzeichnet.

## Data / Evidence Statement

Primaerevidenz umfasst Register (Entscheidungen, Interventionen, Incidents, Evidenz-Manifest), Deploy-Manifeste und -Kette, Logs, Zustandsartefakte, Test- und Acceptance-Evidenz, Freeze-Records und maschinell erzeugte Evaluationsartefakte. Reported State (Anzeigen, Indizes, externe Pakete) und Authoritative State (Datenbanken, Register, Manifeste) werden getrennt gefuehrt. Fehlende Daten werden nicht als Null interpretiert; UNKNOWN, NOT TESTED, NOT VERIFIED, NOT INSTRUMENTED, NOT ESTIMABLE und NOT OBSERVED sind zulaessige Ergebnisse. Live-Datenbanken wurden nie direkt geoeffnet. Personenbezogene Daten und Zugangsdaten sind nicht Teil der Forschungsartefakte; die Biografie des Autors ist Kontext, keine Evidenz.

## Table of Contents (Quelle: S4 Manuskript Core, S5 Appendix Master; keine Seitenzahlen vor dem Layout)

Seitenzahlen fuehrt ausschliesslich die gerenderte PDF-Fassung (LaTeX-generiertes Inhaltsverzeichnis, physische Seiten); die Markdown-Fassung enthaelt keine Seitenzahlen.

- Kapitel I · INTRODUCTION
- Kapitel II · THEORY AND RELATED WORK
- Kapitel III · RESEARCH DESIGN AND METHODOLOGY
- Kapitel IV · FOUNDATIONAL STUDY AND TRANSITION FROM WORK 1
- Kapitel V · AS-BUILT ARTEFACT ARCHITECTURE AT WORK-2 FREEZE
- Kapitel VI · STATE, GOVERNANCE, HUMAN AUTHORITY AND EVIDENCE ARCHITECTURE
- Kapitel VII · LONGITUDINAL PHASE MODEL
- Kapitel VIII · EVALUATION BY EVIDENCE TYPE
- Kapitel IX · FAILURE CASES AND NEGATIVE RESULTS
- Kapitel X · HUMAN-AI, HITL, INTERVENTIONS, REGISTRATION
- Kapitel XI · OPERATIONAL AND BUSINESS OUTCOMES
- Kapitel XII · DISCUSSION
- Kapitel XIII · CANDIDATE DESIGN PRINCIPLES
- Kapitel XIV · LIMITATIONS, CONCLUSION, FUTURE RESEARCH
- Appendix A · Foundational Study / Work 1
- Appendix B · Analysis Decisions / Claim Matrix v2
- Appendix C · Methodology / Timeline
- Appendix D · As-built Artefact Inventory
- Appendix E · Evaluation Method Matrix
- Appendix F · Failure Cases
- Appendix G · RQ Evidence / Traceability
- Appendix H · HITL / Human Decision Evidence
- Appendix I · Interventions / Incidents
- Appendix J · Deploy / Manifest / Acceptance / Chain
- Appendix K · Negative / Null / Not-Estimable Results
- Appendix L · Instrument Registry
- Appendix M · T0/T1 Phase State Table
- Appendix N · Literature / Theory Alignment
- Appendix O · Current Study Freeze
- Appendix P · Reproducibility / Publication Package

## Appendix Overview

[technisches Detail in der nicht öffentlichen Research Edition]

# **HUMAN x AI x REALITY - Longitudinale Folgeuntersuchung des Loewen Codex OS**

**ARBEIT 2 · MANUSKRIFT CORE v1 · Kapitel I-XIV · 2026-09-20**

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL I • INTRODUCTION

## 1.1 Problemstellung

Kleine Organisationen bauen zunehmend eigene Human-AI-Systeme, in denen lokale Sprachmodelle, deterministische Automatisierung, externe Dienste und ein oder zwei Menschen dieselben Geschäftsprozesse tragen. Solche Systeme sind lebendig: sie werden während des Betriebs weiterentwickelt, ihre Anzeigen sind nicht ihr Zustand, und ihre Kontrollmechanismen gelten nur dort, wo sie eingebaut wurden. Die Frage, was ein solches System zu einem gegebenen Zeitpunkt tatsächlich tut - im Unterschied zu dem, was es anzeigt, dokumentiert oder verspricht -, ist damit selbst ein Forschungsproblem. Work 1 (Foundational Study, Freeze 2026-08-22) hat das Artefakt Loewen Codex OS retrospektiv rekonstruiert und zehn Designprinzipien als Hypothesen formuliert; sie hat aber, wie ihr eigener Abschluss-Audit festhält, zentrale Befunde ohne Zahl, Kernzahlen mit schwacher Evidenzklasse und keine Möglichkeit hinterlassen, 'gebaut' von 'gelaufen' zu unterscheiden (Kap. IV).

## 1.2 Forschungsbedarf

Drei Lücken motivieren die vorliegende Arbeit. Erstens fehlt für lebende, vom Betreiber selbst entwickelte Systeme eine Evidenzarchitektur, die Artefaktzustand, Messzustand, Evidenzzustand und Claim-Scope trennt; ohne sie werden Anzeigen zu Befunden und Labels zu Handlungen. Zweitens ist unklar, ob technisch erzwungene Governance in realen, extern wirksamen Aktionspfaden greift - die Work-1-Aussage dazu war qualitativ. Drittens fehlt ein Umgang mit Messinstrumenten, die selbst Teil des untersuchten Systems sind und selbst ausfallen können. Die Literatur zu Design Science (Hevner et al. 2004 [PARTIAL]; Peffers et al. 2007 [PARTIAL]; Gregor & Hevner 2013 [PARTIAL]), Human Oversight (Sterz et al. 2024 [PARTIAL]; Green 2022), Dependability (Avizienis et al. 2004) und Provenienz (Herschel et al. 2017) stellt Begriffe bereit, beschreibt aber nicht den Fall eines lebenden Artefakts, das für die Evaluation eingefroren werden muss (Kap. II).

## 1.3 Work 1 -> Work 2 Motivation

Work 1 forderte die prospektive Prüfung seiner Prinzipien selbst (Anhang F, 7.9). Work 2 setzt deshalb einen T1-Freeze (2026-09-06T22:56:02), eine Präregistrierung für eine Extension-Woche und einen abschliessenden Freeze (2026-09-20T08:46:41) und beobachtet das Artefakt dazwischen. Sie ersetzt Work 1 nicht und mischt keine Korpora: Augustbefund bleibt Augustbefund, Septemberbefund bleibt Septemberbefund (Kap. IV, Appendix A).

## 1.4 Forschungsfragen

- **RQ0** Unter welchen Bedingungen lässt sich der Zustand eines lebenden Human-AI-Artefakts belastbar feststellen und beschreiben?
- **RQ1** Wie belastbar ist State Verification (operational observability) im Artefakt, und welche Evidenzebenen sind dafür nötig?
- **RQ2** Wie funktionieren Governance und Enforcement in realen, extern wirksamen Aktionspfaden?
- **RQ3** Wie ist menschliche Autorität gegenüber Agentenausführung im Artefakt verteilt, und wie gut ist sie beobachtbar (HITL)?

- **RQ4** Wie veraendert sich das Artefakt longitudinal, und wie vollstaendig wird Veraenderung registriert?
- **RQ5** Ist das Artefakt operativ bzw. oekonomisch wirksam?
- **RQ6** Wie treu repraesentieren Anzeigen, Berichte und Instrumente den autoritativen Zustand?

## 1.5 Forschungsziel

Ziel ist eine longitudinale, scope-sensitive Evaluation des Artefakts nach Design-Science-Logik: jede Aussage traegt Population, Zeitfenster, Evidenzklasse und Status (VERIFIED, SCOPED, NOT TESTED, NOT VERIFIED, NOT MEASURED, NOT ESTIMABLE). Nicht Ziel ist der Nachweis operativer, oekonomischer oder rechtlicher Wirksamkeit; RQ5 wird gestellt, um ihre Nichtbeantwortbarkeit mit dem vorliegenden Design zu dokumentieren.

## 1.6 Beitrag

Die Arbeit liefert sechs fallbezogene Beitraege (Kap. XIV.3): longitudinale Evidenz aus einem heterogenen Single-Operator-Artefakt; eine methodische Trennung von Artefakt-, Mess- und Evidenzzustand; Failure Cases, die Instrumentfehler einschliessen; eine fallbasierte Operationalisierung fail-closed Kontrollen; ein Reproduzierbarkeitsgeruest aus append-only Registern, Hash-Ketten und Freeze; und acht Candidate Design Principles zur kuenftigen Validierung. Keiner dieser Beitraege wird als neu im Sinne einer systematischen Literaturbasis beansprucht; die bevorzugten Formulierungen sind extends, operationalizes, combines, illustrates.

## 1.7 Scope

Untersucht wird der Zustand des Artefakts bis zum Freeze 2026-09-20T08:46:41. Ausgeschlossen sind: Post-Freeze-Aenderungen, T2-Ergebnisse (DRAFT), Biografie (Referenz, keine Evidenz), externe Synthesepakete (E1/REPORTED), Realversand von E-Mails (NOT TESTED) und jede rechtliche Bewertung (E1 OPEN).

## 1.8 Aufbau der Arbeit

Teil I (I-III) legt Fragen, Theorie und Design; Teil II (IV-VI) beschreibt Foundational Study, As-built-Architektur und Evidenzarchitektur; Teil III (VII-XI) evaluiert nach Phasen, Evidenzarten, Failure Cases, Human-AI-Verteilung und Outcomes; Teil IV (XII-XIV) diskutiert, formuliert Designkandidaten und schliesst mit Limitations, Conclusion und T2-Grenze. Appendices A-P tragen Register, Tabellen, Hashes und Maschinendaten.

# KAPITEL II • THEORY AND RELATED WORK

Grundlage ist ausschliesslich die S1-Literature-Matrix (45 Quellen; {'PARTIALLY VERIFIED': 15, 'VERIFIED': 25, 'VERIFIED (als Preprint)': 5}). PARTIALLY VERIFIED Quellen werden nur fuer bestaetigte Aspekte zitiert und tragen den Marker [PARTIAL]; Preprints tragen [PREPRINT] und dienen nur der Statusangabe. Wo beim Schreiben eine weitere Quelle noetig schiene, steht LITERATURE GAP.

## 2.1 Design Science Research

Design Science Research (DSR) versteht Wissen als Ergebnis von Bau und Anwendung eines Artefakts und stellt Leitlinien fuer Ausfuehrung und Evaluation bereit (Hevner et al. 2004 [PARTIAL]). Die DSRM-Prozessfolge - Problemidentifikation, Ziele, Design/Entwicklung, Demonstration, Evaluation, Kommunikation - macht Evaluation und Kommunikation zu eigenen Aktivitaeten (Peppers et al. 2007 [PARTIAL]); Work 2 ordnet seine Phasen (Kap. VII) dieser Folge zu. Gregor & Hevner unterscheiden Beitragstypen nach Reife von Problem- und Loesungsdomaene und Abstraktionsstufen; die Candidate Design Principles dieser Arbeit sind danach hoechstens ein Level-2-Beitrag (nascent design theory) (Gregor & Hevner 2013 [PARTIAL]). March & Storey trennen designbasierte Problemloesungsfragen von theoriebasierten Kausalfragen - RQ5 ist eine Kausalfrage, die ein Design-Bau-Zyklus ohne Kontrollbedingung nicht beantwortet (March & Storey 2008). Arnott & Pervans Bewertung von 362 DSR-Arbeiten zeigt systematische Schwaechen bei Evaluation und Theoretisierung und ist der Massstab, an dem Work 2 seine eigenen NOT-ESTIMABLE-Klassen offenlegt (Arnott & Pervan 2012). FEDS liefert die Evaluationsachsen formativ/summativ und artificial/naturalistic, mit denen Kap. VIII die Methoden klassifiziert (Venable et al. 2016 [PARTIAL]).

## 2.2 Action Design Research als ergaenzende Perspektive

ADR verschraenkt Bauen, Intervention und Evaluation und setzt den Forscher in die Organisation (Sein et al. 2011 [PARTIAL]). Die Konstellation von Work 2 - Betreiber = Entwickler = Forscher - aehnelt dem, ohne dass ein vollstaendiger ADR-Zyklus mit Organisationspartnern beansprucht wird; ADR dient als Abgrenzung, nicht als Selbstzuschreibung.

## 2.3 Human-AI Interaction

Amershis 18 Guidelines fuer Human-AI-Interaktion (u.a. klar machen, was das System kann und warum es tat, was es tat) sind der Referenzrahmen, mit dem Preview-vor-Ausfuehrung und Provenienzanzeige als konsistent - nicht als abgeleitet - eingeordnet werden (Amershi et al. 2019 [PARTIAL]). Die Agenten-Literatur beschreibt Agenten ueber die Module Profile, Memory, Planning und Action (Wang et al. 2024); Kap. V zeigt, dass die 'Agenten' des Artefakts nur Profile und teilweise Memory erfuellen.

## 2.4 Automation, Human Oversight und Trust

[technisches Detail in der nicht oeffentlichen Research Edition]

werden nicht als Literaturklassen ausgegeben (Parasuraman et al. 2000 [PARTIAL]). Misuse (Over-reliance), Disuse (Nichtnutzung nach Fehlalarmen) und Abuse strukturieren den Umgang mit Waechtern und Alarmen (Parasuraman & Riley 1997); Complacency und Automation Bias werden als ueberlappende, aufmerksamkeitsbasierte Phaenomene beschrieben (Parasuraman & Manzey 2010), mit Omission- und Commission-Fehlern als experimentell gezeigten Formen (Skitka et al. 1999). Trust ist nach Lee & See auf angemessene, nicht maximale Reliance zu kalibrieren - die Scope-Angaben dieser Arbeit sind ein Kalibrierungsmittel (Lee & See 2004). Bainbridges Ironien der Automation - der Operator bleibt Monitor seltener Ausnahmen und verliert Uebung - beschreiben die Lage der Waechter-Familie des Artefakts (Bainbridge 1983). Fuer Human Oversight nennen Sterz et al. vier Wirksamkeitsbedingungen (kausale Macht, epistemischer Zugang, Selbstkontrolle, Intention) (Sterz et al. 2024 [PARTIAL]); Green zeigt an 40 Policies, dass Oversight-Vorgaben ein falsches Sicherheitsgefuehl erzeugen koennen (Green 2022). Beide sind das Analyseraster fuer Kap. X.

## 2.5 AI Governance

[technisches Detail in der nicht öffentlichen Research Edition]

## 2.6 Reliability Engineering

Die Dependability-Taxonomie liefert die Kette Fault -> Error -> Failure und unterscheidet Erkennung von Maskierung (Avizienis et al. 2004); sie ist die Terminologie der Failure Cases in Kap. IX. Leveson fasst Sicherheit als Kontrollproblem: Unfaelle entstehen aus unzureichender Durchsetzung von Constraints, nicht nur aus Komponentenausfaellen - Gates sind Control Actions, und eine Anzeige, die nicht dem Betrieb entspricht, ist ein fehlerhaftes Prozessmodell des Controllers (Leveson 2011). Gray belegt Administrations- und Operatorfehler als dominante Ausfallursache und fuehrt das Fail-fast-Prinzip ein (Gray 1985 [PARTIAL]). Huang et al. beschreiben Gray Failure und differential observability: die Detektoren eines Systems melden gesund, waehrend Anwendungen leiden (Huang et al. 2017 [PARTIAL]) - die praeziseste Fassung von 'Port offen ist kein gesunder Dienst'. Chaos Engineering definiert Verifikation durch Steady-State-Hypothese, Stoe-rung und Vergleich (Basiri et al. 2016); die Gegenproben dieser Arbeit sind die Testvariante davon. Circuit-Breaker-Platzierung ist nur als Preprint belegt (Montesi & Weber 2016 [PREPRINT]).

## 2.7 Fail-safe, Fail-closed, Default-deny

'Fail-safe defaults' (Entscheidung auf Erlaubnis statt Ausschluss), 'complete mediation' und 'least privilege' stammen aus dem Security-Kontext von Saltzer & Schroeder (Saltzer & Schroeder 1975). Diese Arbeit uebertraegt sie als ANALOGY / DESIGN TRANSFER auf die Freigabe extern wirksamer Aktionen: fail-closed bedeutet hier fallbezogen, dass bei unklarer oder ungueltiger Voraussetzung die Aktion blockiert wird - kein Systemattribut. Reasons Schichtenmodell mit aktiven Fehlern und latenten Bedingungen begruendet Mehrfachevidenz (Reason 2000); Sklets Barrieren-Definition (Funktion, Zuverlaessigkeit, Reaktionszeit) schaerft die Unterscheidung zwischen Waechter und Tuer (Sklet 2006 [PARTIAL]).

## 2.8 Operational Observability

Observability stammt aus der Regelungstheorie: ein Zustand ist observierbar, wenn er aus den Ausgaben rekonstruiert werden kann (Kalman 1960). Der Praktikerbegriff, den Niedermaier et al. in 28 Interviews erheben, ist weiter und bezeichnet Voraussetzungen stabiler Dienste (Niedermaier et al. 2019). Work 2 verwendet 'operational observability' als Faehigkeit, den relevanten technischen und funktionalen Zustand aus verfuegbarer Telemetrie und Evidenz zu rekonstruieren; die Kalman-Definition dient nur als Kontrast: ein statisches ok:true ist ein Ausgang, aus dem der Zustand gerade nicht rekonstruierbar ist. Praktikerliteratur zu Monitoring (Symptom vs Ursache, Black-/White-Box) wird als graue Quelle gefuehrt (Beyer et al. 2016 [PARTIAL]).

## 2.9 Provenance, Auditability, Reproducibility

[technisches Detail in der nicht öffentlichen Research Edition]

## 2.10 Measurement Reliability und Instrumentation Threats

Runeson & Hoest fassen Validität in Fallstudien als construct, internal, external und reliability (Runeson & Hoest 2009); Shadish, Cook & Campbell führen Instrumentation - die Änderung des Messinstruments im Beobachtungszeitraum - als Bedrohung der internen Validität (Shadish et al. 2002). Praeregistrierung trennt Prediction von Postdiction durch zeitgestempelte Festlegung (Nosek et al. 2018); Work 2 unterscheidet danach T0\_POSTHOC und T1\_PROSPEKTIV.

## 2.11 Agent Architectures (Abgrenzung)

Der begutachtete Agenten-Baukasten (Wang et al. 2024) wird nur zur Abgrenzung genutzt: die Loewen-Codex-‘Agenten’ sind Prompt-/Rollen-/Routing-Konstrukte ohne Tool-Calling; Planning- und Action-Module fehlen (LOG-0332). Die vier in Work 1 geführten arXiv-Surveys sind Preprints (Huang et al. 2024 [PREPRINT]; Zhang et al. 2024 [PREPRINT]; Masterman et al. 2024 [PREPRINT]; Xu et al. 2024 [PREPRINT]) und tragen in Work 2 keinen Claim. Qu et al. wird nur in der 2025 publizierten Fassung geführt (Qu et al. 2025 [PARTIAL]) und nur, falls Kap. V lokale Inferenz einordnet.

## 2.12 Analytisches Rahmenmodell

Aus 2.1-2.11 folgt das Raster dieser Arbeit: (a) Evidenzklassen und Claim-Status je Aussage (DSR-Evaluation, Praeregistrierung); (b) Authoritative vs Reported State und operational observability (Gray Failure, STAMP); (c) Governance als Control Actions mit fail-closed Defaults (Design Transfer); (d) Human Authority als beobachtbare Spur, nicht als Label (Oversight-Bedingungen, Automation Bias); (e) Instrument als Teil des Systems (Instrumentation Threat, Dependability); (f) Provenienz und Freeze als Grenze zwischen Evaluation und Weiterentwicklung. Die sieben Ungleichungen aus Kap. XII sind die fallbezogene Form dieses Rasters.

# KAPITEL III • RESEARCH DESIGN AND METHODOLOGY

## 3.1 DSR Study Design

Work 2 ist eine longitudinale Design-Science-Fallstudie an einem bereits existierenden Artefakt. Bau und Evaluation waren im Beobachtungszeitraum verschraenkt (Post-T1-Build 18.-20.09.), die Evaluation folgte der DSRM-Aktivitaet 5 mit eigener Kommunikation (Aktivitaet 6) in Form dieser Monographie und ihrer Register.

## 3.2 Longitudinal N-of-1 Design

Ein System, ein Principal Operator, eine zweite Operatorrolle (Caller) als Nutzer fuer Teilpfade. Keine Kontrollgruppe, keine Randomisierung, keine Vergleichsbedingung ohne Artefakt oder ohne Mensch. Die Zeitreihe reicht von T0 (12.-22.08., historischer Kontext aus Work 1) bis zum Freeze 2026-09-20T08:46:41.

## 3.3 Artefact / Operator Units

Untersuchungseinheiten sind (a) Komponenten des Artefakts (F-3 Inventar: 60 Inventareintraege gesamt, davon 49 IN-SCOPE-Komponenten, 16 Gates, 14 State Sources), (b) Aenderungssaetze (Deploy-Manifeste), (c) Entscheidungen (Registerzeilen mit HITL-Klasse), (d) Vorfaelle (Incidents), (e) Abnahmen (Sidecars). Die Einheiten haben verschiedene Grundgesamtheiten und werden nicht ineinander umgerechnet.

## 3.4 Phasen

T0 (12.-22.08., Work-1-Kontext) - Pre-T1 (23.08.-06.09.; Attribution C09 am 03./05.09.) - T1 ab 2026-09-06T22:56:02 (Baseline-Woche 07.-13.09.) - Extension (praeregistriert 12./13.09., Fenster 14.-18.09., Early Closeout 16./17./18.09.) - Post-T1-Build ab nominalem Cutover 18.09. 20:00 (erster Post-Freeze-Edit im Intervall (19:59:02, 20:07:12], Intervallbreite 8 min 10 s; exakter Zeitpunkt NOT MEASURED; 19:59:02 liegt 58 s vor dem nominalen Cutover 20:00, daher ist ein Edit bis maximal 58 s vor 20:00 nicht ausschliessbar - diese 58 s sind nicht die Intervallbreite) - CURRENT STUDY FREEZE 2026-09-20T08:46:41. Details Kap. VII.

## 3.5 Data Sources

## 3.6 Evidence Classes

Klassen 1-7 (Terminology Register): 1 Primaerartefakt/Register, 2 Observed durch benannte Person, 3 reproduzierbarer Report, 4 Arbeitsdokument, 5 Handoff/Operatorbericht, 6 Interpretation, 7 extern/REPORTED. Jeder Claim der Claim-Matrix (Appendix B, 26 Claims) traegt Klasse und Status.

## 3.7 Maturity Model

Leiter DECLARED -> IMPLEMENTED -> EXECUTED -> FUNCTIONALLY TESTED -> OPERATIONALLY USED -> MEASURED EFFECTIVE, mit Zusatzattributen DEPLOYED, RUNTIME\_VERIFIED, DELIVERED\_VERIFIED, SCOPED\_BROWSER\_ACCEPTANCE. Kein globaler Reifegrad; Reife wird je Komponente ausgewiesen. Der Index vom 11.09. traegt eine andere Skala (REPORTED) und wird nicht harmonisiert.

## 3.8 HITL-P / HITL-R / HITL-B

Projektinterne Klassen registrierter Entscheidungen: P = prospektiv vor der Handlung eingetragen, R = retrospektiv als Label registriert (u.a. Importe aus comms und Tageslauf), B = Betriebs-/Vorrangentscheidung. Anteile werden nur mit N = klassifizierte Zeilen berechnet; unklassifizierte Zeilen sind nicht HITL = 0. Ein R-Label ist kein bestaetigter menschlicher Review.

## 3.9 Evaluation Methods

Neun Methoden (Kap. VIII): Static/Structural, Unit/Integration, Counterfactual, Regression, Runtime, Delivered, Manifest/Chain, Scoped Browser Acceptance, Zero-Write. Nach FEDS sind Tests und Counterfactuals artificial-formative, Runtime/Delivered artificial-summative, Browser-Acceptance naturalistic-summative im benannten Scope.

## 3.10 Claim Analysis Procedure

[technisches Detail in der nicht öffentlichen Research Edition]

## 3.11 Literature Verification

Targeted literature verification and theory alignment (kein systematisches Review): 45 Quellen, Status {'PARTIALLY VERIFIED': 15, 'VERIFIED': 25, 'VERIFIED (als Preprint)': 5}, Search Log je Feld, W1-Errata (9), DP-Alignment. VERIFIED in Work 1 galt nicht automatisch fuer Work 2.

## 3.12 Validity

[technisches Detail in der nicht öffentlichen Research Edition]

## 3.13 Ethics / Privacy / Legal Boundary

Personenbezogene Daten und Zugangsdaten bleiben ausserhalb der Forschungsartefakte (Public/Private-Split in S8). Fremde Bildschirmfotos werden nicht uebernommen. Rechtliche

Wirksamkeit (Werbe-E-Mail, Einwilligung, Aufzeichnung) wird nicht bewertet: E1 OPEN. Der Anwaltsfall (LOG-0412) ist als Vorfall dokumentiert, nicht als Rechtsbefund.

### **3.14 Reproducibility**

Alle Tabellen dieser Arbeit werden von versionierten Generatoren aus Primaerdaten erzeugt; jeder Schritt (F-1..F-6, S1, S3, S4) hat eine SIEGEL-Datei mit sha256. Live-Datenbanken sind nicht teilbar; nach Peng liegt die Arbeit bei 'Code + Teildaten'. Keine retrospektive Praeregistrierung wird behauptet: nur die Extension-Woche war praeregistriert, und ihre Straenge blieben ueberwiegend NOT OBSERVED.

# KAPITEL IV • FOUNDATIONAL STUDY AND TRANSITION FROM WORK 1

[technisches Detail in der nicht öffentlichen Research Edition]

Generator S3-1.0 (2026-09-20). Jede Aussage traegt einen Work-1-Quellenbezug in eckigen Klammern (Work 1 §x = Monographie Private Deep v1.1, Abschnitt; Steuerdatei = FINAL\_MASTERDOKUMENT 13.09.; keine Seitenzahlen im Haupttext). Work-1-Zahlen sind Work-1-Zahlen (Erhebungsstand je Metrik 21.08./12.08./22.08.) und werden nirgends mit Work-2-Zaehlungen verrechnet. Work 1 bleibt unveraendert.

## 4.1 Research Context of Work 1

[technisches Detail in der nicht öffentlichen Research Edition]

## 4.2 Artefact State 22.08.2026

[technisches Detail in der nicht öffentlichen Research Edition]

### 4.3 Key Findings of Work 1

Auf Ergebnisebene beantwortete Work 1 seine Fragen so: Integration gelingt durch pragmatische Kopplung gemeinsamer Datenbanken, JSON-Zustände, lokaler HTTP-Dienste, Reverse Proxy und Task Scheduler, bleibt aber stark host- und pfadabhängig; entscheidend ist die Trennung generativer KI von deterministischer Ausführung (RQ1). Hohe Automatisierung ist bei repetitiven, klar begrenzten Prozessen erreichbar; Grenzen entstehen bei externer Dienstabhängigkeit, Datenqualität, fehlender Ende-zu-Ende-Telemetrie, regulatorischer Aussenwirkung und semantisch definierten Zielzuständen (RQ2). Technisch erzwungene Gates sind wirksamer als Prompt- oder Dokumentationsregeln; Governance wirkt nur, wenn sie systematisch an produktiven Risikopunkten angewandt wird, Opt-in-Gates erzeugen Lücken (RQ3). Aktivität, Autonomie und Wirksamkeit müssen getrennt gemessen werden; die wichtigsten Lernquellen sind Fehlzustände, veraltete Zustände, ungenutzte Schutzmechanismen und Diskrepanzen zwischen Bericht und Realität (RQ4) [Work 1 §6.8]. Drei Reliability-Fallstudien (0-Byte-Datenbankattrappe, VRAM-Thrashing, Encoding/Plattformgrenzen) tragen das Muster 'Anzeige ist nicht Betrieb' [Work 1 §6.12-6.14; §6.7; Appendix D]. Daraus leitete Work 1 zehn Designprinzipien ab - Deterministic Boundaries, Process-Level Autonomy, Default-Deny Governance, End-to-End Observability, Governance follows Exposure, Fresh Knowledge, Global Stop Capability, Separate Activity from Outcome, Versioned Change Control, Human Oversight by Risk - und formulierte sie ausdrücklich als prüfbare Hypothesen für weitere Zyklen [Work 1 §7.2; §7.9].

### 4.4 Key Limitations of Work 1

Work 1 benennt seine Grenzen selbst: Einzelfallnatur, Betreiber als Entwickler, fehlende randomisierte Kontrollgruppen, fehlende Postmaster-Daten, teilweise fehlende Laufzeitlemetrie und fehlende Versionskontrolle; die Ergebnisse sind analytisch, nicht statistisch generalisierbar [Work 1 §3.6; §7.5]. Die Entwicklung war kein vorab registriertes Experiment; Betreiberbeobachtungen werden niedriger gewichtet als Code und Laufzeitdaten, kausale Wirkungen werden nicht behauptet [Work 1 §3.8]. Der Abschluss-Audit der Steuerdatei fügt vier Befunde hinzu, die für Work 2 leitend sind. Erstens trägt der RQ3-Kernbefund ('ein grosser Anteil schreibender Endpunkte besitzt kein Gate') keine Zahl - weder die schreibenden noch die gegateten Endpunkte wurden gezählt, sodass Designprinzip DP3 quantitativ unbelegt ist [Steuerdatei 6.1-6.4]. Zweitens sind die Kernzahlen (19.000 / 1.000 / 48 / 0) in Anhang J durchgängig Evidenz B/C, während die Arbeit selbst für quantitative Kernaussagen Evidenz A oder B fordert [Steuerdatei 6.1-6.4]. Drittens ist die Anonymisierung der Absolutzahlen rechnerisch reversibel, weil exakte Prozentwerte neben geschwächten Absolutwerten stehen (Entscheidung D-4 offen) [Steuerdatei 6.1-6.4]. Viertens weist der Literatur-Endcheck drei bibliografische Korrekturen aus (Titel Arnott/Pervan, Titel Zhang et al., Status Qu et al.) und führt fünf arXiv-Quellen ohne Peer-Review-Status [Steuerdatei 15]. Der Readiness-Status von Work 1 ist 'READY\_FOR\_PATCH, bedingt' - die Entscheidungen D-1 bis D-8 stehen aus, und der empirische Scope bleibt auf den 22.08.2026 eingefroren; September-Evidenz ist ausdrücklich ausgeschlossen [Steuerdatei 12.4; Steuerdatei Kopf/Abschn. 11].

### 4.5 Why a longitudinal second study was necessary

Work 1 forderte die prospektive Prüfung selbst: Anhang F legt ein Experimentprotokoll mit vorab fixierter Hypothese, Population, Kontrollbedingung, Einfrieren von Konfiguration und Rohdaten und der Reihenfolge 'erst Reliability und Safety instrumentieren, dann Conversion optimieren' vor; Work-1-Abschnitt 7.9 versteht die zehn Designprinzipien als Hypothesen für Folgezyklen; Anhang E listet die Messlücken [Work 1 Appendix F; §7.9; Appendix E; §7.6]. Drei Gründe machten die zweite Studie notwendig. (1) Retrospektive Rekonstruktion kann 'gebaut' und 'gelaufen' nicht unterscheiden - genau das war der Befund vom 03.09.2026, als zwölf vollständig gebaute Subsysteme ohne ein einziges Artefakt gefunden wurden (LOG-0355; Work-2-Zeitstrahl). (2) Der Work-1-Governance-Befund war qualitativ; ob technische Gates in realen Aktionspfaden greifen, lässt sich nur an beobachteten Aktionen prüfen. (3) Die Wirksamkeitsfrage war in Work 1 durch die Attributionslücke ('0 zuordenbare Termine') gar nicht adressierbar; Work 2 hat diese Lücke als Attributionsdefekt identifiziert (Crosswalk-Zeile 5: 1 von 5 Buchungen nachträglich attribuiert), ohne die Wirksamkeit damit zu messen.

4.6 Transition from retrospective to prospective / longitudinal evidence

Work 2 setzt einen T1-Freeze (06.09.2026 22:56:02), eine Praeregistrierung fuer die Extension-Woche (12./13.09.), einen dokumentierten Cutover zum Post-T1-Build (18.09. 20:00) und einen abschliessenden CURRENT STUDY FREEZE (20.09.2026 08:46:41) (F-2 Zeitstrahl). Der Uebergang gelang nur teilweise: Die praeregistrierten Straenge A, B, D und F kamen im Fenster nicht zur Messung (Early Closeout 18.09.; Register AUF-041..044 offen), Strang C blieb partiell, Strang E lieferte Observability-Befunde. Die belastbarsten technischen Zahlen von Work 2 - 17 Deploy-Manifeste, 32 Counterfactuals, 8 Acceptance-Sidecars - entstanden POST-T1 und liegen damit ausserhalb des praeregistrierten Fensters (F-4; F-6 Limitations B). Work 2 ist deshalb im Ergebnis staerker retrospektiv als geplant; der Unterschied zu Work 1 liegt weniger in der Prospektivitaet als in der Instrumentierung: Evidenzklassen je Aussage, Maturity-Leiter je Komponente, append-only Register mit Hash-Ketten, versionierte Messinstrumente und ein Freeze-Record mit Hash.

4.7 Research delta Work 1 -> Work 2

[technisches Detail in der nicht öffentlichen Research Edition]

Quellen- und Claim-Check (Crosswalk-Integration)

Satzhinweis: Breite Tabellen werden in dieser Lesefassung ausschliesslich im A4-Hochformat als zweispaltige Datensaeetze bzw. kompakte Longtables gesetzt. Dies ist eine reine Layouttransformation; Inhalt, Werte und Evidenzstatus bleiben unveraendert.

[technisches Detail in der nicht öffentlichen Research Edition]

|                                                                 |          |
|-----------------------------------------------------------------|----------|
| [technisches Detail in der nicht öffentlichen Research Edition] |          |
| Status                                                          | REFINED  |
| Verwendung in Kap. IV                                           | 4.3, 4.7 |

Crosswalk 2 - RQ2 (W1): Hohe Automatisierung bei repetitiven Prozessen (Mail L4, Content L5, Voice L4);

|                       |                                                   |
|-----------------------|---------------------------------------------------|
| W1-Quelle             | 6.8; Anhang C L0-L5-Matrix                        |
| Status                | CONFIRMED (Grenzen) / NOT RE-TESTED (Level-Skala) |
| Verwendung in Kap. IV | 4.3                                               |

**Crosswalk 3 - RQ3 (W1): Technisch erzwungene Gates wirksamer als Prompt-/Doku-Regeln; Governance wirkt n**

---

|                              |                                                              |
|------------------------------|--------------------------------------------------------------|
| <b>W1-Quelle</b>             | 6.8, 6.4; Steuerdatei 6.3 M4: Kernbefund traegt KEINE ZAHL ( |
| <b>Status</b>                | CONFIRMED (Muster) / REFINED (Inhomogenitaet konkret belegt) |
| <b>Verwendung in Kap. IV</b> | 4.3, 4.4, 4.7                                                |

---

**Crosswalk 4 - RQ4 (W1): Aktivitaet, Autonomie und Wirksamkeit muessen getrennt gemessen werden; Lernquel**

---

|                              |                                                        |
|------------------------------|--------------------------------------------------------|
| <b>W1-Quelle</b>             | 6.8, 6.7, Anhang D                                     |
| <b>Status</b>                | CONFIRMED / REFINED (Instrument als eigener Fehlerort) |
| <b>Verwendung in Kap. IV</b> | 4.3                                                    |

---

**Crosswalk 5 - Mail-Funnel: ~19.365 Sends, ~1.047 Klicks, 48 Antworten (~0,25 %), 0 zuordenbare Termine**

---

|                              |                                                              |
|------------------------------|--------------------------------------------------------------|
| <b>W1-Quelle</b>             | 6.2; Anhang A RQ4; Steuerdatei 6.2 M1, 6.4 (Anonymisierung r |
| <b>Status</b>                | REFINED (Attribution) / NOT COMPARABLE (Zaehlfenster)        |
| <b>Verwendung in Kap. IV</b> | 4.2, 4.5, 4.7                                                |

---

**Crosswalk 6 - Voice: 1.565 Wahlversuche, 1.564 nicht erreicht, 0 Termine**

---

|                              |               |
|------------------------------|---------------|
| <b>W1-Quelle</b>             | 6.3; Anhang A |
| <b>Status</b>                | REFINED       |
| <b>Verwendung in Kap. IV</b> | 4.2           |

---

**Crosswalk 7 - Content Publishing: Versuche 496 / Erfolge 456 / Fehler 41 / 91,7 %**

---

|                              |                                                              |
|------------------------------|--------------------------------------------------------------|
| <b>W1-Quelle</b>             | 6.x; Steuerdatei 6.3 M2: rechnerisch inkonsistent, NOT ESTAB |
| <b>Status</b>                | NOT RE-TESTED                                                |
| <b>Verwendung in Kap. IV</b> | 4.2, 4.7                                                     |

---

**Crosswalk 8 - Governance-Befund: fehlendes Default-Deny, kein globaler Kill Switch, confirm-Gates, Sandb**

---

|                                                                                |                        |
|--------------------------------------------------------------------------------|------------------------|
| <b>W1-Quelle</b>                                                               | Anhang A RQ3; Anhang C |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                        |
| <b>Verwendung in Kap. IV</b>                                                   | 4.7                    |

---

**Crosswalk 9 - Reliability-Fallstudien I-III: 0-Byte-DB-Attrappe, VRAM-Thrashing, Encoding/cp1252**

---

|                              |                                                       |
|------------------------------|-------------------------------------------------------|
| <b>W1-Quelle</b>             | 6.12-6.14                                             |
| <b>Status</b>                | CONFIRMED (Muster 'Anzeige $\neq$ Betrieb') / REFINED |
| <b>Verwendung in Kap. IV</b> | 4.3                                                   |

---

---

**Crosswalk 10 - DP1 Deterministic Boundaries**

---

|                              |                       |
|------------------------------|-----------------------|
| <b>W1-Quelle</b>             | 7.2                   |
| <b>Status</b>                | CONFIRMED (fallweise) |
| <b>Verwendung in Kap. IV</b> | 4.3                   |

---

---

**Crosswalk 11 - DP2 Process-Level Autonomy**

---

|                              |           |
|------------------------------|-----------|
| <b>W1-Quelle</b>             | 7.2       |
| <b>Status</b>                | CONFIRMED |
| <b>Verwendung in Kap. IV</b> | 4.3       |

---

---

**Crosswalk 12 - DP3 Default-Deny Governance**

---

|                              |                                          |
|------------------------------|------------------------------------------|
| <b>W1-Quelle</b>             | 7.2; Steuerdatei M4 quantitativ unbelegt |
| <b>Status</b>                | REFINED                                  |
| <b>Verwendung in Kap. IV</b> | 4.7                                      |

---

---

**Crosswalk 13 - DP4 End-to-End Observability (Port/HTTP 200/Queue reichen nicht)**

---

|                              |                                                        |
|------------------------------|--------------------------------------------------------|
| <b>W1-Quelle</b>             | 7.2                                                    |
| <b>Status</b>                | CONFIRMED / REFINED (mehr Ebenen $\neq$ mehr Wahrheit) |
| <b>Verwendung in Kap. IV</b> | 4.3                                                    |

---

---

**Crosswalk 14 - DP5 Governance follows Exposure**

---

|                              |                                                    |
|------------------------------|----------------------------------------------------|
| <b>W1-Quelle</b>             | 7.2                                                |
| <b>Status</b>                | CONFIRMED (Richtung) / NOT RE-TESTED (Rechtsseite) |
| <b>Verwendung in Kap. IV</b> | 4.7                                                |

---

---

**Crosswalk 15 - DP6 Fresh Knowledge (RAG Freshness)**

---

|                              |                               |
|------------------------------|-------------------------------|
| <b>W1-Quelle</b>             | 7.2                           |
| <b>Status</b>                | NOT RE-TESTED (kein W2-Claim) |
| <b>Verwendung in Kap. IV</b> | 4.7                           |

---

---

**Crosswalk 16 - DP7 Global Stop Capability**

---

|                              |               |
|------------------------------|---------------|
| <b>W1-Quelle</b>             | 7.2           |
| <b>Status</b>                | NOT RE-TESTED |
| <b>Verwendung in Kap. IV</b> | 4.7           |

---

**Crosswalk 17 - DP8 Separate Activity from Outcome**

---

|                              |           |
|------------------------------|-----------|
| <b>W1-Quelle</b>             | 7.2       |
| <b>Status</b>                | CONFIRMED |
| <b>Verwendung in Kap. IV</b> | 4.3       |

---

**Crosswalk 18 - DP9 Versioned Change Control (kein Git; Kopien + Logbuch)**

---

|                              |                          |
|------------------------------|--------------------------|
| <b>W1-Quelle</b>             | 7.2; LOG-0332            |
| <b>Status</b>                | REFINED (Harness gebaut) |
| <b>Verwendung in Kap. IV</b> | 4.7                      |

---

**Crosswalk 19 - DP10 Human Oversight by Risk**

---

|                              |                                  |
|------------------------------|----------------------------------|
| <b>W1-Quelle</b>             | 7.2                              |
| <b>Status</b>                | REFINED (Provenienz statt Label) |
| <b>Verwendung in Kap. IV</b> | 4.3                              |

---

**Crosswalk 20 - Evidenzregister W1: 122 Belege (65+29+2+18+8), Klassen A-E**

---

|                              |                         |
|------------------------------|-------------------------|
| <b>W1-Quelle</b>             | Anhang; Steuerdatei 6.1 |
| <b>Status</b>                | NOT COMPARABLE          |
| <b>Verwendung in Kap. IV</b> | -                       |

---

**Crosswalk 21 - QA-Zahlen im Logbuch ('1.420 Faele', '916 Simulationspruefungen')**

---

|                              |                               |
|------------------------------|-------------------------------|
| <b>W1-Quelle</b>             | LOG-0332 (nicht zitierfaehig) |
| <b>Status</b>                | REFINED (erstmalig belegbar)  |
| <b>Verwendung in Kap. IV</b> | -                             |

---

**Crosswalk 22 - Anonymisierung der Absolutzahlen ist rechnerisch reversibel (D-4 offen)**

---

|                              |                 |
|------------------------------|-----------------|
| <b>W1-Quelle</b>             | Steuerdatei 6.4 |
| <b>Status</b>                | NOT RE-TESTED   |
| <b>Verwendung in Kap. IV</b> | 4.4             |

---

**Crosswalk 23 - 'Doctoral-level research' auf dem W1-Cover = gemessener Widerspruch (Steuerdatei 16)**

---

|                              |                |
|------------------------------|----------------|
| <b>W1-Quelle</b>             | Steuerdatei 16 |
| <b>Status</b>                | REFINED        |
| <b>Verwendung in Kap. IV</b> | 4.1            |

---

REFINED-Zeilen (9): RQ1, RQ3, RQ4, Mail-Funnel, Voice, Governance-Befund, DP3, DP9, DP10, Doctoral-Cover (Statusverteilung des Crosswalks: {'REFINED': 9, 'CONFIRMED': 9, 'NOT': 5}). NOT RE-TESTED / NOT COMPARABLE (5): RQ2 Level-Skala, Content-Quote, DP6, DP7, Evidenzregister, QA-Zahlen, Anonymisierung (teils REFINED/NOT RE-TESTED gemischt - siehe Crosswalk-JSON). Keine Zeile CONTRADICTED.

## Regeln, die dieses Kapitel einhaelt

- Work-1-Zahlen nur als Work-1-Zahlen mit Erhebungsstand je Metrik
- keine September-Evidenz in Work-1-Aussagen; Work-2-Bezeuge nur als Verweis auf

[technisches Detail in der nicht öffentlichen Research Edition]

Zitat, nicht als Korrektur

- keine Zitation aus der Biografie; kein Bezug auf Masterpack v14
- kein Grad, kein Peer Review behauptet

## Work-1-Referenzhashes

SHA-256-Werte sind aus Lesbarkeitsgruenden in zwei 32-Zeichen-Zeilen umbrochen.

- **Monographie Private Deep v1.1 (MD-Extraktion 03.09.)**

[technisches Detail in der nicht öffentlichen Research Edition]

- **Private Deep PDF (29.08., Deckblatt 'Version 1.0')**

[technisches Detail in der nicht öffentlichen Research Edition]

Upload U1, Register 20.09.

- **Steuerdatei FINAL\_MASTERDOKUMENT 13.09.**

[technisches Detail in der nicht öffentlichen Research Edition]

- **Public PDF WEBSITE\_COVERED 13.09. (= Webroot)**

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL V • AS-BUILT ARTEFACT ARCHITECTURE AT WORK-2 FREEZE

Beschrieben wird der Zustand am 2026-09-20T08:46:41, as-built, nicht desired. Grundlage ist das F-3-Inventar (Appendix D) mit F3-Erratum.

## 5.1 Architekturebenen

[technisches Detail in der nicht öffentlichen Research Edition]

## 5.2 Komponenten und Maturity

[technisches Detail in der nicht öffentlichen Research Edition]

## 5.3 State Sources und Entities

14 State Sources (authoritative/reported) und 9 Entitäten mit Nicht-Gleichsetzungen (Firma != Lead != Kontaktpunkt; Operator-Authentifizierung != Rollenregistry, OPERATOR-B login canonicalization (case variant) als Canonicalization-Konflikt erhalten). Der Lead liegt in sechs Ablagen ohne führende Quelle (LOG-0359); companies.id ist die führende Identität des Post-T1-Baus.

## 5.4 Caller Workbench, Contact Points, Manual Mail

[technisches Detail in der nicht öffentlichen Research Edition]

## 5.5 manual\_send\_requests (F3-Erratum)

Der Idempotenz-Mechanismus `manual_send_requests` ist als Code/Schema IMPLEMENTED (CREATE TABLE IF NOT EXISTS beim ersten Send, Tests auf Wegwerf-DB gruen). Sein Live-Zustand ist NOT OBSERVED / NOT INSTANTIATED - die Tabelle existierte im Freeze nicht, weil nie gesendet wurde -, seine Ausfuehrung NOT TESTED. Er ist kein persistenter Live-State-Store am Freeze.

## 5.6 Harness und Runtime

`deploy_manifest.py` erzeugt je Aenderungssatz ein Manifest mit sha256 vor/nach, Tests, Counterfactuals, Regressionen, Runtime-Checks, Verification Model A-E und Kette (prev/self); `build_view.py` liest sie read-only; Acceptance-Sidecars sind separate append-only Dateien. Der Harness ist Werkzeug, kein Agent. Seine Semantik aenderte sich ab LOG-0501 (Kap. VIII).

## 5.7 Public Representation

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL VI • STATE, GOVERNANCE, HUMAN AUTHORITY AND EVIDENCE ARCHITECTURE

## 6.1 Authoritative vs Reported State

Authoritative State liegt in Backend-Datenbank, Registern und Manifesten; Reported State in Frontend, Cockpits, Index und externen Paketen. Das Frontend repräsentiert Zustand, es besitzt ihn nicht (F02: Parser-Konflikte; Kanban-localStorage vs CRM). In PROV-Vokabular: Manifest = Entity, Deploy = Activity, Operator und Harness = Agents (Moreau & Missier (W3C) 2013).

## 6.2 Governance Gates

16 beschriebene Gates (Send-Tor, Suppression adressweit, Identitäts-Riegel, Ownership, Preview-Invalidation, Sender-Binding, Company-Link-Filter, Perimeter-Allowlists, Reload-Tore u.a.). Ihr Charakter ist fail-closed im fallbezogenen Sinn (Terminology Register); ihre Abdeckung ist nicht homogen. Nach Sklet sind sie aktive technische Barrieren nur dort, wo eine Reaktion folgt - ein Wächter, der nur meldet, ist keine Barriere (Sklet 2006 [PARTIAL]).

## 6.3 Human Authority vs Agent Execution

Menschliche Autorität ist beobachtbar, wo sie eine eigene Spur hinterlässt: HITL-P-Zeilen, Acceptance-Sidecars, human\_decision\_ref in Manifesten, Freeze-Entscheidung. Der Dev-Agent ist gated und wurde nie benutzt; der Harness ist read-only; Preview steht vor Send. Die 'Agenten' des Artefakts führen keine Tool-Calls aus. Gleichzeitig laufen die ältesten Ketten (Cold-Mail, IG-Publishing) modellfrei ohne Vorgangsfreigabe - Human Authority ist im neuen Caller-Pfad implementiert, nicht systemweit.

## 6.4 Verification Model A-E

[technisches Detail in der nicht öffentlichen Research Edition]

## 6.5 Evidence Architecture und Claim Scope

Jeder Claim trägt Evidenzklasse, Population, Zeitfenster, Maturity und Status; Zusammenfassungen dürfen den Scope nicht verlieren. Die Arbeit verwendet hierfür die fallbezogene Formulierung 'evidence scope travels with the claim' (DP-C4) - keine etablierte Terminologie, sondern eine case-derived synthesis aus Provenienz, Präregistrierung und Trust-Kalibrierung (Herschel et al. 2017; Nosek et al. 2018; Lee & See 2004).

## 6.6 Nicht-Gleichsetzungen

- Frontend != Authoritative State
- Label != Human Action
- Deploy != Intervention
- Manifest State != Reliability
- Acceptance != Effectiveness
- Eingefroren != Erfasst

# KAPITEL VII • LONGITUDINAL PHASE MODEL

Grundlage: F-2 Timeline (45 Zeilen, Appendix C). Phasen strikt getrennt; keine Post-Freeze-Evidenz.

## 7.1 Historischer Kontext und T0

Work 1 friert am 2026-08-22 ein; T0-Fenster 12.-22.08. Die Work-1-Zahlen bleiben Work-1-Zahlen (Kap. IV).

## 7.2 Pre-T1 (23.08.-06.09.)

Inventur der zwölf nie gelaufenen Subsysteme (03.09.), Forschungsinstrumentierung (Register, Evidenz-Manifest, Transkriptsicherung), Attributionskette C09 (03./05.09., PRE-T1: 1 von 5 Buchungen nachträglich attribuiert), DR-Gates, Research Protocol v0.1-0.3.

## 7.3 T1 (ab 2026-09-06T22:56:02)

T1\_START mit ABSCHLUSS\_SIEGEL (Erhebung 22:55:58; Ordnername 22:55:38 ist keine Quelle). Baseline-Woche 07.-13.09. mit Ramp-up der zweiten Operatorrolle, drei Re-Freezes am 10.09. (18:27:28, 19:48:04, 22:40:15 WINDOW\_B\_3\_START; Datei-Hash-Gegenprobe ausstehend). HITL-Instrument endet 10.09. 20:40:51.

## 7.4 Extension (14.-18.09.)

[technisches Detail in der nicht öffentlichen Research Edition]

## 7.5 Post-T1-Build (18.-20.09.)

Nominaler Cutover 18.09. 20:00; erster Post-Freeze-Edit im Intervall (19:59:02, 20:07:12], Intervallbreite 8 min 10 s; exakter Zeitpunkt NOT MEASURED; 19:59:02 liegt 58 s vor dem nominalen Cutover 20:00, daher ist ein Edit bis maximal 58 s vor 20:00 nicht ausschliessbar - diese 58 s sind nicht die Intervallbreite. LOG-0484..0493 ohne Manifest (nur LOG-/Observed-Evidenz), LOG-0494..0506 mit 17 FINAL-Manifesten (P-DEP). Drei registrierte Interventionen (INT-0030..0032).

## 7.6 CURRENT STUDY FREEZE

[technisches Detail in der nicht öffentlichen Research Edition]

|                          |                                                                          |
|--------------------------|--------------------------------------------------------------------------|
| Study Freeze             | 2026-09-20T08:46:41                                                      |
| Record-Hash              | 5c1c6244...                                                              |
| Deploy-/Acceptance-Kette | 26 Zeilen = 17 Manifest + 8 Acceptance + 1 Freeze                        |
| Freeze-Register          | 15 Finding-Eintraege; 13 nicht geschlossen; OF-I / OF-K RESOLVED         |
| Abgrenzung               | getrennt von 18 Index-known_conflicts, 56 Incidents und 16 Failure Cases |
| Folgephase               | T2 DRAFT; keine Empirie in Work 2 nach dem Freeze                        |

# KAPITEL VIII • EVALUATION BY EVIDENCE TYPE

Population P-DEP = 17 FINAL-Manifeste ({'FAILED': 5, 'VERIFIED': 7, 'WARN': 5}). Alle Zahlen im Numeric Register; Detailmatrix Appendix E. Der Abschnitt fragt je Methode, was sie zeigt und was nicht.

## Static / Structural Checks

[technisches Detail in der nicht öffentlichen Research Edition]

## Unit / Integration Tests

[technisches Detail in der nicht öffentlichen Research Edition]

## Counterfactual Tests

**Ergebnis.** 32 PASS / 0 FAIL - jede Gegenprobe biss **Zeigt:** Alt-Fassung hatte die Luecke; Neu-Fassung schliesst sie; Test ist diskriminierend **Zeigt nicht:** Business-Wirkung; ungetestete Pfade **Maturity:** FUNCTIONALLY TESTED (Designaenderung). **Grenze:** -

## Regression Tests

**Ergebnis.** 126 PASS / 1 FAIL (FAIL: dev1\_test U4/X9 in 0634 durch transienten DRIFT einer Fixture - behoben durch Fixture-Haertung, s. F06) **Zeigt:** keine Regression in den geprueften Suiten beobachtet **Zeigt nicht:** 'keine Regression existiert' **Maturity:** FUNCTIONALLY TESTED (Erhalt). **Grenze:** Suiten decken nur getestete Pfade

## Runtime Verification

[technisches Detail in der nicht öffentlichen Research Edition]

# Delivered Verification

[technisches Detail in der nicht öffentlichen Research Edition]

## Manifest / Chain

**Ergebnis.** 7 VERIFIED, 5 WARN (LIVE\_DB\_MTIME\_DURING\_TESTS 4x, NOT\_VERIFIABLE\_CHECKS 1x), 5 FAILED (TEST\_FAIL 3, RUNTIME\_FAIL 2, REGRESSION\_FAIL 1, NOT\_VERIFIABLE 3);  
Gruende: { 'TEST\_FAIL': 3, 'NOT\_VERIFIABLE\_CHECKS': 4, 'RUNTIME\_FAIL': 2, 'LIVE\_DB\_MTIME\_DURING': 5, 'REGRESSION\_FAIL': 1 }; Self-Verify in allen finalisierten Laufen PASS; Kette 26 Zeilen ohne Bruch **Zeigt:** Deploy-Identität und -Historie fuer gelistete Dateien **Zeigt nicht:** Systemqualität ('VERIFIED' ist kein Rating) **Maturity:** Zusatzattribut DEPLOYED. **Grenze:** Vollständigkeit der Liste

## Scoped Browser Acceptance

**Ergebnis.** alle Sidecars PASS oder PASS\_SPLIT; letzter (LOG0506 Zeile 2): 11 PASS, 1 NOT\_VERIFIED, 4 NOT\_TESTED **Zeigt:** UI-Verhalten in benannten Fällen **Zeigt nicht:** globale Verifikation; Real-Send; Operational Use **Maturity:** SCOPED\_BROWSER\_ACCEPTANCE. **Grenze:** Skript-Sidecar (OF-H)

## Zero-Write

**Ergebnis.** Delta 0 in allen vier (sends\_max\_id 52083 == 52083; Tabelle nicht vorhanden) **Zeigt:** im beobachteten Acceptance-Fenster wurden keine entsprechenden Writes registriert **Zeigt nicht:** 'Preview schreibt grundsätzlich nie'; Preview-Anzahl (NOT\_INSTRUMENTED) **Maturity:** OBSERVED\_ZERO (Messung). **Grenze:** Fenster kurz, ein Operator

## Zusammenfassung und Measurement Change

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL IX · FAILURE CASES AND NEGATIVE RESULTS

Failure Cases sind der empirische Kern von RQ6. Terminologie nach der Dependability-Taxonomie (Avizienis et al. 2004). Sechzehn Faelle (Appendix F) werden hier nach Failure-Semantik gruppiert; je Gruppe wird ein Hauptfall ausgeführt, die uebrigen werden verwiesen. Root Cause bleibt UNKNOWN, wo sie nicht diagnostiziert wurde.

## Object Failure - Hauptfall F01: HTTP 200 bei nicht funktionsfaehiger UI

[technisches Detail in der nicht öffentlichen Research Edition]

## Integration Failure - Hauptfall F02: Backend-Schema vs Frontend-Parser

**Kontext.** Cockpit las inv.items, Backend lieferte inv.catalog (LOG-0407); Frontend-Feldnamen netto/brutto vs summe/betrag (LOG-0307 (10)) **Erwartet.** Frontend zeigt Backend-Daten **Beobachtet.** leere/falsche Anzeige ('[object Object]', '-') **Evidenz.** HAUPTLOG LOG-0407, LOG-0307; Index-Konflikte (Klasse 1). **Root Cause.** Vertrag zwischen Schreiber und Leser nicht getestet **Intervention / Post-State.** Vertragstests (TestClient + jsdom, LOG-0398 Integrationspruefung) / Integrationspruefung 0 Abweichungen (LOG-0398) **Maturity.** FUNCTIONALLY TESTED **Lehre.** Schema != Rendering; Vertragstest noetig **Weitere Faelle dieser Gruppe:** F01 (Appendix F).

## State Failure - Hauptfall F14: manual\_send\_requests implementiert, nicht instanziiert

**Kontext.** CREATE TABLE IF NOT EXISTS beim ersten Send; F-3 nannte es State Source **Erwartet.** Live-Tabelle vorhanden **Beobachtet.** Tabelle im Freeze nicht vorhanden (Zero-Write 07:19/08:35) **Evidenz.** BLOCK31B\_ZERO\_WRITE\_\*.json; F3-Erratum (Klasse 1/2). **Root Cause.** kein Send erfolgt **Intervention / Post-State.** F3-Erratum append-only / IMPLEMENTED / NOT INSTANTIATED / NOT TESTED **Maturity.** IMPLEMENTED **Lehre.** Schema-Mechanismus != Live-State **Weitere Faelle dieser Gruppe:** F06 (Appendix F).

[technisches Detail in der nicht öffentlichen Research Edition]

[technisches Detail in der nicht öffentlichen Research Edition]

## **Governance Failure - Hauptfall F13: caller\_demo\_link Governance-Inkonsistenz**

**Kontext.** Demo-Link-Pfad nutzt company\_ref\_aus\_pool ohne LINK\_METHODEN\_MAIL-Filter **Erwartet.** gleicher Standard auf beiden Mail-Pfaden **Beobachtet.** zwei Standards im selben Layer **Evidenz.** LOG-0506 E; F-3 Konflikt (Klasse 1). **Root Cause.** Haertung nur im neuen Modul **Intervention / Post-State.** keine (Freeze) / OPEN SEND-RISK **Maturity.** - **Lehre.** Governance-Mechanismen sind nicht automatisch systemweit konsistent **Weitere Faelle dieser Gruppe:** F11 (Appendix F).

## **Instrumentation Failure - Hauptfall F03: FAILED Deploys**

[technisches Detail in der nicht öffentlichen Research Edition]

## **Registration Failure - Hauptfall F16: Deploy vs Intervention Registration Gap**

[technisches Detail in der nicht öffentlichen Research Edition]

## **Incident-Population und negative Ergebnisse**

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL X • HUMAN-AI, HITL, INTERVENTIONS, REGISTRATION

## 10.1 HITL-Register

[technisches Detail in der nicht öffentlichen Research Edition]

## 10.2 Label ist nicht Handlung

Die 543 HITL-R-Zeilen sind REGISTERED LABELS aus Import-Pfaden (331 comms, 212 Tageslauf); confirmed human retrospective reviews sind NOT ESTIMABLE. Gegen die Wirksamkeitsbedingungen von Sterz et al. (Sterz et al. 2024 [PARTIAL]) erfuehlt ein Label keine nachweisbar; HITL-P-Zeilen und Sidecars erfuehlen kausale Macht und epistemischen Zugang fuer ihren Pfad. Das Artefakt implementierte damit mehrere explizite Human-Authority-Punkte; 'effective human oversight system-wide' wird nicht behauptet.

## 10.3 Acceptance als menschliche Abnahme

[technisches Detail in der nicht öffentlichen Research Edition]

## 10.4 Interventionen und Registrierung

134 Interventionsregister-Zeilen (IDs mehrfach; nicht 134 bestaetigte Interventionen) ({'T0\_POSTHOC': 124, 'T1\_PROSPEKTIV': 10}; Status {'null': 108, 'verworfen': 22, 'IMPLEMENTED\_NOT\_VALIDATED': 1, 'REGISTERED\_BEFORE\_EDIT\_NOT\_YET\_IMPLEMENT': 3}); nach dem Cutover nur ['INT-0030', 'INT-0031', 'INT-0032']. 17 Deploys sind nicht 17 Forschungsinterventionen: die Deploy-Kette ist vollstaendiger als das Register, Registration Coverage ist NOT ESTIMABLE, weil die Grundgesamtheit 'alle Aenderungen' unbeobachtet ist. Die Drift ist ein Methodenbefund; eine Alternative ist die Registrierungsregel selbst (Pflicht nur fuer P0-A). Praeregistrierung trennt Prediction von Postdiction (Nosek et al. 2018) - T0\_POSTHOC-Zeilen sind Postdiction.

# KAPITEL XI • OPERATIONAL AND BUSINESS OUTCOMES

[technisches Detail in der nicht öffentlichen Research Edition]

# KAPITEL XII • DISCUSSION

Basis: F-5 (Interpretation), S1 (Einordnung), F-6 Literature Patch (append-only). Je Kernbefund: Observed - Interpretation - Literature Alignment - Counterevidence - Alternative Explanation - Boundary Condition. Es werden keine Results neu erzeugt.

## RQ0 - Beschreibbarkeit und Artefaktgrenze (ANSWERED DESCRIPTIVELY)

[technisches Detail in der nicht öffentlichen Research Edition]

## RQ1 - Observability / State Verification (ANSWERED DESCRIPTIVELY)

**Observed.** Runtime 142 PASS / 2 FAIL / 4 NOT VERIFIABLE; Delivered alle PASS; Kette 26 Zeilen; Zero-Write Delta 0; Failure Cases F01 (HTTP 200 ohne Funktion), F05 (PASS + unhandledRejection), F06 (false/transienter DRIFT), F14 (Schema ohne Live-State).

**Interpretation.** Verifizierbarkeit erforderte im Fall mehrere Evidenzebenen (Hash -> Manifest -> Kette -> Runtime -> Delivered -> Browser). Jede einzelne Ebene wurde mindestens einmal durch einen Gegenbefund als unzureichend gezeigt: HTTP-Status (F01), Teststatus (F05), Disk-Status (F06), Schema (F14). Authoritative State (Backend/DB) und Reported State (Frontend, Cockpit, Index) muessen getrennt gefuehrt werden - das Frontend repraesentiert Zustand, es besitzt ihn nicht.

**Literature Alignment.** (Huang et al. 2017 [PARTIAL]; Kalman 1960; Niedermaier et al. 2019; Avizienis et al. 2004)

**Counterevidence.** Mehr Ebenen erzeugten selbst Fehler (F04 LIVE\_DB\_MTIME, F06 Pointer-Semantik, OF-J). NOT\_VERIFIABLE-Checks (4) sind ehrliche Luecken, keine Belege.

**Alternative Explanation.** Die Gegenbefunde koennten Artefakte einer noch unreifen Harness sein (Semantik geaendert ab LOG-0501) statt Eigenschaften des Systems.

**Boundary Condition.** Fallbezogen; keine Aussage ueber Verifizierbarkeit anderer Stacks; Fenster 19.-20.09. fuer Harness-Evidenz.

## RQ2 - Governance / Enforcement (PARTIALLY ANSWERED)

[technisches Detail in der nicht öffentlichen Research Edition]

## RQ3 - Human-AI / HITL (PARTIALLY ANSWERED)

[technisches Detail in der nicht öffentlichen Research Edition]

## RQ4 - Longitudinal Change (PARTIALLY ANSWERED)

**Observed.** 134 Interventionsregister-Zeilen (IDs mehrfach; nicht 134 bestaetigte Interventionen) ({‘T0\_POSTHOC’: 124, ‘T1\_PROSPEKTIV’: 10}), 17 Manifeste POST-T1, 3 Interventionen nach Cu-tover (INT-0030..0032), Amendments, Sidecars - fuenf Evidenzpopulationen mit verschiedener Vollstaendigkeit; F16.

**Interpretation.** Operational Development und Research Registration drifteten auseinander: die Deploy-Kette ist vollstaendiger als das Interventionsregister. 17 Deploys sind nicht 17 Forschungsinterventionen, und Registration Coverage ist NOT ESTIMABLE, weil die Grundgesamtheit ‘alle Aenderungen’ selbst nicht beobachtet ist. Diese Drift ist ein Methodenbefund, kein Nebenprodukt.

**Literature Alignment.** (Nosek et al. 2018; Haeberlen et al. 2007 [PARTIAL]; Crosby & Wallach 2009)

**Counterevidence.** Die Deploy-Kette existiert erst ab 19.09. (Harness-Bau); fuer 07.-18.09. gibt es

keinen vergleichbaren Nenner; HAUPTLOG-Luecke LOG-0483 -> 0484.

**Alternative Explanation.** Die Drift koennte durch die Registrierungsregel (Pflicht nur fuer P0-A) erklart sein, nicht durch Unterlassung - dann waere sie Design, nicht Fehler.

**Boundary Condition.** Kurze Zeitreihe (T1 06.09. - Freeze 20.09.), keine Nachregistrierung.

## RQ5 - Effectiveness (NOT MEASURED)

**Observed.** Instrumentiert nur: 906 Sends (Tagesbericht 16.09.), 10 offene Antworten, 1/5 attribuierte Buchung (PRE-T1), retainers 0 Zeilen, MRR-Provenienz offen. Extension F NOT OBSERVED. C13 NOT PERMISSIBLE.

**Interpretation.** Technische Funktionsfaehigkeit darf nicht zu Business-Wirksamkeit hochgestuft werden: es gibt keine Kontrollgruppe, keine kausale Attribution (die Attributionskette ist erst seit LOG-0411 E2E belegt und traegt genau eine Buchung), und keine Outcome-Metrik mit Nenner. Sends, Replies, Clicks, Bookings sind deskriptive Zaehlungen; Klicks sind zu grossen Teilen Maschinen (LOG-0427).

**Literature Alignment.** (March & Storey 2008)

**Counterevidence.** -

**Alternative Explanation.** Jeder beobachtete Geschaefstwert koennte durch Aktivitaeten ausserhalb des Artefakts (Telefon, Empfehlung, Bestandskunden) entstanden sein.

**Boundary Condition.** RQ5 bleibt offen; keine Ersatzantwort ueber Proxys.

## RQ6 - Representation Fidelity (ANSWERED DESCRIPTIVELY)

[technisches Detail in der nicht öffentlichen Research Edition]

## Sieben Ungleichungen

[technisches Detail in der nicht öffentlichen Research Edition]

- **IMPLEMENTATION != BUSINESS EFFECT** - 16 Gates beschrieben, 906 Sends gezaehlt (Report 16.09.), 0 Retainer-Zeilen (Snapshot 03.09., NOT RE-MEASURED), 1 attribuierte Buchung PRE-T1 -> C13 NOT PERMISSIBLE; kein Proxy als Effekt

## Instrumentfehler und Measurement Change

Drei von fuenf FAILED-Manifesten sind Instrumentfehler; FAILED ist ein Evidenzzustand, kein Qualitaetsrating, und Instrumentfehler reduzieren dennoch die Aussagekraft der Evaluation. Der Semantikwechsel ab LOG-0501 ist eine Instrumentation-Bedrohung (Shadish et al. 2002); P-DEP wird nicht als homogene Messreihe gelesen. Die Gegenprobe ist die Testvariante der Chaos-Engineering-Logik (Basiri et al. 2016); ihre Staerke ist auf die Testmatrix begrenzt.

## Fail-closed als Design Transfer

Fail-safe defaults, complete mediation und least privilege (Saltzer & Schroeder 1975) wurden als Analogie auf Aktionsfreigaben uebertragen. Im untersuchten Fall war fail-closed Verhalten fuer mehrere risikorelevante Pfade ein nuetzliches Muster; caller\_demo\_link, die nicht implementierte P0-G-Capability, Booking-Divergenz und E1 bleiben Gegenbeispiele. Kein fail-safe Gesamtsystem.

## Neue Widersprueche

- DP-C5 (mehr Ebenen) und DP-C7 (Instrument als Failure Mode) stehen in Spannung: die Evidenz fuer C5 stammt aus derselben Harness, deren Fehler C7 begruenden - kein unabhaengiger Beleg fuer den Nettonutzen.
- DP-C6 wird durch die modellfreien Automaten (Cold-Mail, IG-Publishing ohne Vorgangsfreigabe, LOG-0332) im selben System konterkariert - 'Human Authority' gilt fuer den neuen Caller-Pfad, nicht fuer die aeltesten, taeglich wirkenden Ketten.
- RQ2 'fail-closed bewaehrt' und F13 'Nachbarpfad ungehaertet' sind derselbe Sachverhalt aus zwei Richtungen; wer nur die Browserfaelle zitiert, ueberschaetzt die Governance.
- Die Extension-Woche liefert das prospektive Design der Studie, aber keine prospektiven Daten; die belastbarsten Zahlen (Manifeste, Counterfactuals) sind POST-T1 und damit ausserhalb des praeregistrierten Fensters - die Studie ist im Ergebnis staerker retrospektiv als geplant.

# KAPITEL XIII · CANDIDATE DESIGN PRINCIPLES

Alle acht Prinzipien sind CANDIDATE (single-system, N-of-1). SUPPORTED bedeutet Literaturanschluss, nicht empirische Validierung. Formulierungen: operationalizes / combines / consistent with.

## DP-C1 · AUTHORITATIVE STATE SHOULD REMAIN SERVER-SIDE FOR CONSEQUENTIAL ACTIONS

[technisches Detail in der nicht öffentlichen Research Edition]

## DP-C2 · PREVIEW AND EXECUTION SHOULD BE SEPARATE STATES

**Status.** CANDIDATE - Literature Alignment: PARTIALLY ALIGNED. **Empirical Support.** Preview Zero-Write (Delta 0 in vier Tabellen), final\_gate erst im Send, Preview Invalidation (preview\_token, Browser PASS), Sender Binding (sender\_ref, SENDER UNBOUND) **Counterevidence.** Preview garantiert keine Fehlerfreiheit: F07 zeigte den Identitätsfehler IM Preview; Real Send NOT TESTED, also ist die Trennung nur bis zur Preview-Seite belegt **Boundary.** Manual-Mail-Modul; Preview-Anzahl NOT INSTRUMENTED **Alternative Explanation.** Zero-Write könnte am kurzen Fenster mit einem Operator liegen **Literaturanschluss.** L26 (fail-safe defaults), L09 (Guidelines: make clear what/why), L13 (calibrated reliance) **Formulierung.** 'consistent with fail-safe defaults; the preview-invalidation mechanism is an engineering operationalization, not a theoretical contribution'

## DP-C3 · IDENTITY SHOULD BE BOUND MORE STRONGLY THAN HEURISTIC MATCHES BEFORE EXTERNAL ACTION

[technisches Detail in der nicht öffentlichen Research Edition]

variant) Canonicalization-Konflikt; sechs Lead-Ablagen ohne fuehrende Quelle (LOG-0359) **Boundary.** gilt fuer den Mail-Pfad des Caller-Moduls; keine allgemeine Identitaetstheorie **Alternative Explanation.** Die beobachteten Fehlbindungen koennten Datenqualitaetsprobleme (Scraper-Fragmente) sein, nicht Heuristikfehler **Literaturanschluss.** L26 (complete mediation, least privilege), L32 (Provenienz der Bindung) **Formulierung.** 'the case combines provenance display with fail-closed defaults for weak entity links'

## DP-C4 • EVIDENCE SCOPE SHOULD TRAVEL WITH THE CLAIM

**Status.** CANDIDATE - Literature Alignment: CASE-DERIVED SYNTHESIS / PROJECT TERMINOLOGY (NO DIRECT SOURCE FOUND FOR THE EXACT FORMULATION). **Empirical Support.** SCOPED\_BROWSER\_ACCEPTANCE mit expliziten NOT VERIFIED / NOT TESTED; runtime vs delivered als getrennte Attribute; Manifest-States mit Gruenden; F-1 Claim-Matrix mit Statusspalte **Counterevidence.** Zusammenfassungen verloren Scope trotzdem: F-3 nannte manual send\_requests als State Source (Erratum); Index 11.09. traegt eine andere Maturity-Skala; Logbuch-QA-Zahlen ohne Report **Boundary.** gilt fuer die hier gebauten Artefakte; ob Scope beim Uebergang in Fliesstext erhalten bleibt, ist erst in F-6/Manuskript pruefbar **Alternative Explanation.** Scope-Erhalt koennte Ergebnis der Generator-Disziplin sein, nicht des Prinzips **Literaturanschluss.** L32/L33 (Provenienz auf Claim-Ebene), L13 (Kalibrierung), L39 (Prediction/Postdiction-Trennung), L03 (Beitragstyp) **Formulierung.** 'the study operationalizes claim-level provenance; the label is the study's own'

## DP-C5 • VERIFICATION SHOULD USE MULTIPLE INDEPENDENT EVIDENCE LAYERS

**Status.** CANDIDATE - Literature Alignment: SUPPORTED BY PRIOR LITERATURE (mit Gegenbe-fund). **Empirical Support.** hash / manifest / chain / runtime / delivered / browser / counterfactual; jede Ebene fing mindestens einen Fall, den eine andere nicht fing (F01 vs Delivered; F05 vs Runtime; F14 vs Tests) **Counterevidence.** Mehr Instrumentierung erzeugte neue Instrumentierungs-fehler: 3/5 FAILED = Instrument (F03), LIVE\_DB\_MTIME (F04), false DRIFT (F06), OF-J; Meas-urement Change ab LOG-0501 **Boundary.** Ebenen sind nicht vollstaendig unabhaengig (dieselbe Harness, derselbe Operator); Nutzen vs Komplexitaet nicht gemessen **Alternative Explanation.** Der Mehrwert koennte an der Fehlerklasse liegen (Frontend-Faelle), nicht an der Ebenenzahl **Literaturanschluss.** L27 (Defense in Depth), L23 (differential observability), L20 (Erkennung vs Maskierung), L24 (Chaos/Gegenprobe) **Formulierung.** 'the findings are consistent with defence-in-depth; the study adds that each layer introduced its own instrument faults'

## DP-C6 • HUMAN AUTHORITY SHOULD BE DISTINCT FROM AGENT EXECUTION

[technisches Detail in der nicht öffentlichen Research Edition]

## DP-C7 · MEASUREMENT INSTRUMENTS SHOULD BE VERSIONED AND TREATED AS PART OF THE RESEARCH SYSTEM

**Status.** CANDIDATE - Literature Alignment: SUPPORTED BY PRIOR LITERATURE. **Empirical Support.** LOG-0501 Semantikwechsel, OF-J, OF-K, LIVE\_DB\_MTIME, TestClient-Failures, ODR instrument\_version je Zeile, Generator-Versionen F1-F5 **Counterevidence.** Versionierung verhinderte die Drift nicht, sie machte sie nur sichtbar; Alt-QA-Zahlen ohne Testreport bleiben nicht zitierfähig **Boundary.** gilt fuer die im Freeze gelisteten Instrumente; Instrumente ausserhalb (Tagessberichte 14.-18.09.) sind nicht versioniert **Alternative Explanation.** Instrumentfehler koennen auf die Bauzeit (eine Woche) zurueckgehen und mit Reife verschwinden **Literaturanschluss.** L38 (Instrumentation Threat), L37 (Reliability in Case Studies), L34 (Reproducibility), L20 (Fault im Messsystem) **Formulierung.** 'the study applies the instrumentation threat (Shadish et al. 2002) to a self-built deploy harness'

## DP-C8 · FREEZES SHOULD SEPARATE EVALUATION FROM CONTINUOUS DEVELOPMENT

**Status.** CANDIDATE - Literature Alignment: PARTIALLY ALIGNED. **Empirical Support.** CURRENT STUDY FREEZE 08:46:41 mit Record-Hash, append-only Register, 15 Finding-Eintraege im Freeze-Register (13 zum Study Freeze nicht geschlossen; OF-I und OF-K RESOLVED als Historie; getrennt von 18 Index-known\_conflicts, 56 Incidents und 16 Failure Cases) erhalten, T2 DRAFT, F-1..F-5 read-only **Counterevidence.** Freeze eliminiert keinen Bias: T1-Freeze-Abdeckung war lueckenhaft (Webroot ausserhalb), Re-Freezes 10.09. dreimal, Early Closeout 18.09.; 'Eingefroren != Erfasst' **Boundary.** Freeze schuetzt die Evaluation des eingefrorenen Zustands, nicht die Vollstaendigkeit des Zustands **Alternative Explanation.** Die Trennung koennte auch durch Ressourcenmangel (kein Bau moeglich) entstanden sein **Literaturanschluss.** L39 (zeitgestempelte Praeregistrierung), L34 (Reproducibility), L35/L36 (tamper-evident logs), L07 (Evaluationszeitpunkt) **Formulierung.** 'the case combines preregistration timestamps with hash-sealed artefact freezes; no direct literature on freezing a live artefact for evaluation was found'

# KAPITEL XIV • LIMITATIONS, CONCLUSION, FUTURE RESEARCH

F-6 ist die Claim-Schranke dieses Kapitels.

## 14.1 Limitations

### A DESIGN

- N-of-1: ein System, ein Artefakt-Stack, eine organisatorische Umgebung
- single principal operator; eine zweite Operatorrolle (Caller) nur fuer Teilpfade
- keine Kontrollgruppe, keine Randomisierung, keine Vergleichsbedingung ohne Artefakt oder ohne Mensch
- Evaluationsmethoden vom Autor gewaehlt und gebaut (Selektion durch Testfokus)

### B TEMPORAL

- T1-Phase kurz (06.09.-20.09.); Baseline-Woche und Extension ohne Strangmessung
- die staerkste technische Evidenz (17 Manifeste, 32 Counterfactuals, Browser-Acceptance) ist POST-T1 (19.-20.09.) und liegt ausserhalb des praeregistrierten Fensters
- Measurement Change ab LOG-0501: P-DEP nicht unter identischer Harness-Semantik erhoben
- HAUPTLOG-Luecke 13.-18.09.; HITL-Instrument endet 10.09.

### C INSTRUMENTATION

- HITL Coverage: N=579 klassifiziert von 3.224; 2.645 unklassifiziert; confirmed human reviews NOT ESTIMABLE
- Registration Coverage NOT ESTIMABLE (Grundgesamtheit der Aenderungen unbeobachtet)
- Failure Rate NOT ESTIMABLE (kein Exposure-Nenner; Incident-Dubletten)
- Preview Count NOT INSTRUMENTED; Business Effect NOT MEASURED; MRR NOT ESTIMABLE (Provenienzkonflikt)
- Runtime NOT\_VERIFYABLE-Checks (4) sind Luecken, keine Belege; LIVE\_DB\_MTIME (5 Manifeste) = Messkontamination

### D ROLE / BIAS

- dieselbe Person entwickelte, betrieb, priorisierte, testete teilweise, interpretierte und nahm Browserfaelle ab
- Kontrollen: append-only Register mit Hash-Ketten, maschinell erzeugte Tabellen aus Primaerdaten, Counterfactuals mit Gegenprobe, Freeze mit Record-Hash, explizit gefuehrte Counter-evidence - sie reduzieren das Bias-Risiko, eliminieren es nicht
- kein externer Reviewer hat Primaerdaten gesehen; Gutachter-Ruecklaeufe (ChatGPT, Masterpack v14) sind E1/REPORTED

## E SYSTEM

[technisches Detail in der nicht öffentlichen Research Edition]

## F LITERATURE

- targeted verification and theory alignment, kein systematisches Review, kein PRISMA
- 45 Quellen: 25 VERIFIED, 15 PARTIALLY VERIFIED (Publisher-Seiten nicht abrufbar; bleiben PARTIAL), 5 Preprints (separat, in W2 nicht als Beleg)
- NO DIRECT SOURCE: Health-Check-Semantik, Event Sourcing als Forschungsevidenz, Freeze eines lebenden Artefakts, Entity-Resolution-Schwellen
- Literatur wurde ausschliesslich zur Einordnung verwendet; sie ersetzt keine Messung und hebt keinen Status (NOT TESTED/NOT ESTIMABLE) an

### Autor-/Operatorbias (Hervorhebung)

Dieselbe Person entwickelte, betrieb, priorisierte, testete teilweise, interpretierte und nahm Browserfäelle ab. Kontrollen: append-only Logs, maschinell erzeugte Tabellen, Counterfactuals mit Gegenprobe, Freeze, explizit geführte Counterevidence. Diese reduzieren das Bias-Risiko, eliminieren es aber nicht.

## 14.2 Final RQ Answers

- **RQ0 (ANSWERED DESCRIPTIVELY).** Das Artefakt liess sich im eingefrorenen Scope als technisch-organisatorisches Human-AI-System beschreiben, ausführen und evaluieren; die Artefaktgrenze musste dabei selbst hergestellt werden (Freeze-Manifest, Harness-Listen, Inventar sind nicht deckungsgleich). Grenze: kein 'abgeschlossenes Gesamtsystem'; heterogene Reifegrade; 15 Finding-Einträge im Freeze-Register (13 nicht geschlossen, 2 RESOLVED)
- **RQ1 (ANSWERED DESCRIPTIVELY).** Operational observability erforderte im Fall mehrere Evidenzebenen; jede einzelne Ebene wurde durch mindestens einen Gegenbefund als allein unzureichend gezeigt, und die Ebenen erzeugten selbst Instrumentfehler. Grenze: fallbezogen; Harness-Semantik nicht homogen
- **RQ2 (PARTIALLY ANSWERED).** Fail-closed Kontrollen fuer extern wirksame Aktionen konnten in einzelnen Pfaden operationalisiert und geprüft werden (Tests, Runtime, 11 Browserfäelle, 32 Gegenproben); sie waren nicht systemweit homogen (caller\_demo\_link, Booking-Divergenz, P0-G als nicht implementierte Capability). Grenze: Fäelle, keine Grundgesamtheit der Aktionspfade; E1 OPEN
- **RQ3 (PARTIALLY ANSWERED).** Das Artefakt implementierte mehrere explizite Human-Authority-Punkte (HITL-P 18, 8 Acceptance-Sidecars, human\_decision\_ref); die 543 HITL-R sind registrierte Labels, keine bestaetigten Reviews, und die HITL-Abdeckung endet am 10.09. Grenze: confirmed reviews NOT ESTIMABLE; keine Vergleichsbedingung; Automation-Bias-Effekt nicht gemessen
- **RQ4 (PARTIALLY ANSWERED).** Operative Entwicklung und Forschungsregistrierung drifteten auseinander (17 Deploys vs 3 registrierte Interventionen nach Cutover); die Drift ist ein Methodenbefund, ihre Groesse ist NOT ESTIMABLE. Grenze: kurze Zeitreihe; Deploy-Kette erst ab 19.09.

- **RQ5 (NOT MEASURED / NOT ANSWERABLE WITH CURRENT DESIGN).** Operative bzw. oekonomische Wirksamkeit konnte mit dem vorliegenden Design nicht belastbar gemessen werden; das ist ein Ergebnis des Messdesigns (keine Kontrollbedingung, keine Outcome-Metrik mit Nenner, Attribution erst seit LOG-0411), nicht ein Befund 'keine Wirkung'. Grenze: 906 Sends / 10 offene Antworten / 1 von 5 Buchungen attribuiert (PRE-T1) / 0 Retainer-Zeilen (Snapshot 03.09., NOT RE-MEASURED at freeze) nur deskriptiv; MRR NOT ESTIMABLE
- **RQ6 (ANSWERED DESCRIPTIVELY).** Repraesentationstreue erwies sich als Eigenschaft der ganzen Kette Real State -> Observation -> Instrument -> Output; Fehlermodi betreffen Messinstrumente, Tests, Frontend/Backend-Vertraege, Identity Resolution, Registrierung und Verification-Semantik, nicht nur das Artefakt. Grenze: Failure Rate NOT ESTIMABLE; nur belegte Faelle F01-F16

## 14.3 Contributions

[technisches Detail in der nicht öffentlichen Research Edition]

## 14.4 Conclusion

### 1 What was studied?

Untersucht wurde das Loewen Codex OS - ein selbstgebautes, heterogenes Human-AI-Operator-System eines Einzelunternehmers - in einer longitudinalen Folgestudie (Work 2) zur retrospektiven Foundational Study (Work 1, Freeze 22.08.2026). Beobachtet wurden die Phasen T0, Pre-T1, T1 (ab 06.09.), Extension (14.-18.09., Early Closeout), Post-T1-Build (18.-20.09.) und der CURRENT STUDY FREEZE (2026-09-20T08:46:41). Forschungsfragen RQ0-RQ6 zielten auf Beschreibbarkeit, operational observability, Governance/Enforcement, Human-AI-Zusammenarbeit, longitudinale Veraenderung, Wirksamkeit und Repraesentationstreue.

## 2 What was actually demonstrated?

Die Studie zeigt, dass das Loewen Codex OS im untersuchten Zeitraum als technisch-organisatorisches Human-AI-Artefakt implementiert, ausgeführt und in mehreren funktionalen, Governance- und Verifikationsdimensionen evaluiert werden konnte. Belegt sind: 17 finalisierte Deploy-Manifeste (7 VERIFIED / 5 WARN / 5 FAILED - ein Evidenzzustand, kein Qualitätscore), 52 Test-PASS bei 3 FAIL, 32 Counterfactuals, die in den geprüften Fällen zwischen altem und neuem Verhalten diskriminierten, Runtime- und Delivered-Verifikation gelisteter Dateien, Zero-Write-Delta 0 im Acceptance-Fenster und 11 real im Browser abgenommene Fälle im Manual-Mail-Scope (SCOPED\_BROWSER\_ACCEPTANCE). Das Artefakt implementierte mehrere explizite Human-Authority-Punkte (HITL-P 18, 8 Acceptance-Sidecars). Maximal gestützter Reifegrad: FUNCTIONALLY TESTED mit den Attributen DEPLOYED, RUNTIME\_VERIFIED, DELIVERED\_VERIFIED und scope-spezifischer Browser-Abnahme.

## 3 What failed / remained uncertain?

[technisches Detail in der nicht öffentlichen Research Edition]

## 4 What can be learned from the case?

Erstens: Anzeige ist nicht Betrieb, PASS ist nicht fehlerfrei, Entity Match ist nicht Identität, Preview ist nicht Send, Deployed ist nicht Operationally Used, Browser Accepted ist nicht Measured Effective, Implementation ist nicht Business Effect - jede dieser Ungleichungen hat im Fall einen belegten Vorfall. Zweitens: das Messinstrument ist Teil des Forschungssystems und selbst Fehlerquelle; ohne Versionierung und Freeze verschmelzen Evaluation und Weiterentwicklung. Drittens: fail-closed Kontrollen wirkten dort, wo sie eingebaut waren, und nirgends sonst - ein Gate-Katalog ist keine Konsistenzgarantie. Viertens: Human Oversight, die nur als Label existiert, erfüllt keine der in der Literatur genannten Wirksamkeitsbedingungen; beobachtbar ist Autorität nur, wo sie eine eigene Spur hinterlässt. Diese Einsichten sind als acht Candidate Design Principles formuliert - bekannte Prinzipien in neuem Kontext bzw. fallbezogene Kombinationen, nicht validierte Universalprinzipien.

## 5 What should be tested next?

Der methodische Fortschritt von Work 1 (retrospektive Artefakt- und Prozessrekonstruktion) zu Work 2 (longitudinale, scope-sensitive Evaluation mit prospektiven Elementen, Failure Semantics, Measurement Reliability, Human Decision Provenance, Scoped Acceptance und Freeze Boundary) schließt die Work-1-Lücken nicht vollständig: der RQ3-Kernbefund trägt in beiden Arbeiten keine Grundgesamtheit, und die Wirksamkeitsfrage bleibt offen. T2 bleibt DRAFT. Zu prüfen wären: Erhalt der funktionalen und Governance-Eigenschaften unter Skalierung (zweiter Operator, mehr Agenten, höheres Volumen, Reliability under load), Realversand nach E1-Klärung mit instrumentierter Zustellung, Multi-Agent-Governance und ein Wirkungsdesign mit Kontrollbedingung für RQ5. Keine T2-Ergebnisse liegen vor.

## Maximaler Claim (geprüft)

Die Studie zeigt, dass das Loewen Codex OS im untersuchten Zeitraum als technisch-organisatorisches Human-AI-Artefakt implementiert, ausgeführt und in mehreren funktionalen, Governance- und Verifikationsdimensionen evaluiert werden konnte. Die Ergebnisse dokumentieren sowohl funktionierende fail-closed Mechanismen in benannten Pfaden als auch technische Failure Cases, Messinstrumentfehler, Governance-Inhomogenität und substantielle Evidenzlücken. Die Untersuchung belegt keine operative, ökonomische, rechtliche oder generalisierbare

Wirksamkeit. (Abweichung zum Arbeitsentwurf: 'in benannten Pfaden' ergaenzt, weil C-Matrix nur pfadbezogene Governance-Claims traegt.)

## **14.5 Future Research**

Zu pruefen: Erhalt funktionaler und Governance-Eigenschaften unter Skalierung (zweiter Operator, mehr Agenten, hoeheres Volumen, Reliability under load); Realversand nach E1-Klaerung mit instrumentierter Zustellung; Multi-Agent-Governance und Human Control; ein Wirkungsdesign mit Kontrollbedingung fuer RQ5; Root Cause fuer OF-B und OF-J; Grundgesamtheit schreibender Endpunkte fuer den RQ3-Befund beider Arbeiten.

## **14.6 T2 Boundary**

T2 bleibt DRAFT. Keine Ergebnisse, keine rueckwirkende Prospektivitaet. Operative Aenderungen nach dem Freeze werden als POST-FREEZE / E1-E2-Override / T2 getrennt gefuehrt und gehoeren nicht zu dieser Arbeit.

# LITERATURVERZEICHNIS

[technisches Detail in der nicht öffentlichen Research Edition]

[technisches Detail in der nicht öffentlichen Research Edition]

[technisches Detail in der nicht öffentlichen Research Edition]

- [L28] Sklet, S. (2006). Safety barriers: Definition, classification, and performance. *Journal of Loss Prevention in the Process Industries*, 19(5), 494-506 Identifier: 10.1016/j.jlp.2005.12.004 (UNCONFIRMED). [Status S1: PARTIALLY VERIFIED - DOI/Heft unbestaetigt]
- [L15] Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Roensberg, A., Meinel, P., & Langer, M. (2024). On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. *ACM FAccT '24* Identifier: 10.1145/3630106.3659051 (via arXiv Related DOI). [Status S1: PARTIALLY VERIFIED - Seiten unbestaetigt]
- [L17] NIST (Tabassi, E.) (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 Identifier: 10.6028/NIST.AI.100-1.
- [L07] Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, 25(1), 77-89 Identifier: 10.1057/ejis.2014.36. [Status S1: PARTIALLY VERIFIED - Publisher-Seite leer; Achsen aus Vorwissen]
- [L40] Wang, L., et al. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345 Identifier: 10.1007/s11704-024-40231-1.
- [L45] Xu, J., et al. (2024). On-Device Language Models: A Comprehensive Review. *arXiv:2409.00088 (v2)* Identifier: arXiv:2409.00088. [PREPRINT - nicht begutachtet]
- [L43] Zhang, Z., et al. (2024). A Survey on the Memory Mechanism of Large Language Model based Agents. *arXiv:2404.13501 (v1)* Identifier: arXiv:2404.13501. [PREPRINT - nicht begutachtet]

# ARBEIT 2 · APPENDICES MASTER

## v1 · 2026-09-20

Navigations- und Integrationsartefakt. Primaer-Appendix-Dateien (F-1..F-6, S1, S3, S4) werden per Hash referenziert, nicht neu erzeugt. Zahlen tragen die metric\_id des S4 Numeric Register oder den Vermerk APPENDIX-ONLY. Generator S5-1.0 (2026-09-20). ARBEIT 2 NICHT FINAL.

### Master Index

#### Appendix 1 - A

|                                                                                |                                                   |
|--------------------------------------------------------------------------------|---------------------------------------------------|
| <b>title</b>                                                                   | Foundational Study / Work 1                       |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                   |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                   |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                   |
| <b>machine_readable</b>                                                        | yes                                               |
| <b>public_class</b>                                                            | PUBLIC_OK (Crosswalk) / PRIVATE (W1 Private Deep) |
| <b>claim_scope</b>                                                             | W1-Zahlen nur als W1; Freeze 22.08.               |
| <b>primary_or_derived</b>                                                      | DERIVED                                           |
| <b>freeze_relation</b>                                                         | vor W2-Freeze                                     |

#### Appendix 2 - B

|                                                                                |                                      |
|--------------------------------------------------------------------------------|--------------------------------------|
| <b>title</b>                                                                   | Analysis Decisions / Claim Matrix v2 |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                      |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                      |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                      |
| <b>machine_readable</b>                                                        | yes                                  |
| <b>public_class</b>                                                            | PUBLIC_OK                            |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                      |
| <b>primary_or_derived</b>                                                      | DERIVED                              |
| <b>freeze_relation</b>                                                         | post-freeze analysis                 |

#### Appendix 3 - C

|                                                                                |                                                                                           |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| <b>title</b>                                                                   | Methodology / Timeline                                                                    |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                           |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                           |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                           |
| <b>machine_readable</b>                                                        | yes                                                                                       |
| <b>public_class</b>                                                            | PUBLIC_OK                                                                                 |
| <b>claim_scope</b>                                                             | 45 Zeilen; Cutover-Intervall (19:59:02, 20:07:12], <= 58 s vor 20:00 nicht ausschliessbar |
| <b>primary_or_derived</b>                                                      | DERIVED                                                                                   |
| <b>freeze_relation</b>                                                         | post-freeze                                                                               |

## Appendix 4 - D

|                                                                                |                                                                                                                              |
|--------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>title</b>                                                                   | As-built Artefact Inventory                                                                                                  |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                                                              |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                                                              |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                                                              |
| <b>machine_readable</b>                                                        | yes                                                                                                                          |
| <b>public_class</b>                                                            | PUBLIC_ABSTRACT                                                                                                              |
| <b>claim_scope</b>                                                             | 60 Eintraege gesamt / 49 IN SCOPE/16 Gates/14 State Sources; manual_send_requests<br>IMPLEMENTED/NOT INSTANTIATED/NOT TESTED |
| <b>primary_or_derived</b>                                                      | DERIVED                                                                                                                      |
| <b>freeze_relation</b>                                                         | post-freeze                                                                                                                  |

## Appendix 5 - E

|                                                                                |                                        |
|--------------------------------------------------------------------------------|----------------------------------------|
| <b>title</b>                                                                   | Evaluation Method Matrix               |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                        |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                        |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                        |
| <b>machine_readable</b>                                                        | yes                                    |
| <b>public_class</b>                                                            | PUBLIC_OK                              |
| <b>claim_scope</b>                                                             | 9 Methoden; keine PASS-Summe als Score |
| <b>primary_or_derived</b>                                                      | DERIVED                                |
| <b>freeze_relation</b>                                                         | post-freeze                            |

## Appendix 6 - F

|                                                                                |                                           |
|--------------------------------------------------------------------------------|-------------------------------------------|
| <b>title</b>                                                                   | Failure Cases                             |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                           |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                           |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                           |
| <b>machine_readable</b>                                                        | yes                                       |
| <b>public_class</b>                                                            | PUBLIC_ABSTRACT (Pseudonymisierung S8)    |
| <b>claim_scope</b>                                                             | 16 Faelle; Instrument vs Produkt getrennt |
| <b>primary_or_derived</b>                                                      | DERIVED                                   |
| <b>freeze_relation</b>                                                         | post-freeze                               |

## Appendix 7 - G

|                                                                                |                            |
|--------------------------------------------------------------------------------|----------------------------|
| <b>title</b>                                                                   | RQ Evidence / Traceability |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                            |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                            |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                            |
| <b>machine_readable</b>                                                        | yes                        |
| <b>public_class</b>                                                            | PUBLIC_OK                  |
| <b>claim_scope</b>                                                             | RQ5 NOT MEASURED           |
| <b>primary_or_derived</b>                                                      | DERIVED                    |
| <b>freeze_relation</b>                                                         | post-freeze                |

## Appendix 8 - H

|                                                                                |                                |
|--------------------------------------------------------------------------------|--------------------------------|
| <b>title</b>                                                                   | HITL / Human Decision Evidence |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                |
| <b>machine_readable</b>                                                        | yes                            |
| <b>public_class</b>                                                            | PUBLIC_OK                      |

|                           |                                            |
|---------------------------|--------------------------------------------|
| <b>claim_scope</b>        | 543 R = Labels; drei Populationen getrennt |
| <b>primary_or_derived</b> | DERIVED                                    |
| <b>freeze_relation</b>    | post-freeze                                |

## Appendix 9 - I

|                                                                                |                                               |
|--------------------------------------------------------------------------------|-----------------------------------------------|
| <b>title</b>                                                                   | Interventions / Incidents                     |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                               |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                               |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                               |
| <b>machine_readable</b>                                                        | yes                                           |
| <b>public_class</b>                                                            | PRIVATE_ONLY (Register) / PUBLIC_OK (Summary) |
| <b>claim_scope</b>                                                             | 134 / 56 Count; keine Rate                    |
| <b>primary_or_derived</b>                                                      | PRIMARY (Register) + DERIVED                  |
| <b>freeze_relation</b>                                                         | Freeze-Stand                                  |

## Appendix 10 - J

|                                                                                |                                                 |
|--------------------------------------------------------------------------------|-------------------------------------------------|
| <b>title</b>                                                                   | Deploy / Manifest / Acceptance / Chain          |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <b>machine_readable</b>                                                        | partial                                         |
| <b>public_class</b>                                                            | REVIEWER_ONLY (Manifeste) / PUBLIC_OK (Summary) |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <b>primary_or_derived</b>                                                      | PRIMARY + DERIVED                               |
| <b>freeze_relation</b>                                                         | Freeze-Stand                                    |

## Appendix 11 - K

|                                                                                |                                         |
|--------------------------------------------------------------------------------|-----------------------------------------|
| <b>title</b>                                                                   | Negative / Null / Not-Estimable Results |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                         |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                         |
| <b>hash</b>                                                                    | see SIEGEL_S5                           |
| <b>machine_readable</b>                                                        | yes                                     |
| <b>public_class</b>                                                            | PUBLIC_OK                               |
| <b>claim_scope</b>                                                             | vollstaendig, nichts entfernt           |
| <b>primary_or_derived</b>                                                      | DERIVED                                 |
| <b>freeze_relation</b>                                                         | post-freeze                             |

## Appendix 12 - L

|                                                                                |                                    |
|--------------------------------------------------------------------------------|------------------------------------|
| <b>title</b>                                                                   | Instrument Registry                |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                    |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                    |
| <b>hash</b>                                                                    | see SIEGEL_S5                      |
| <b>machine_readable</b>                                                        | yes                                |
| <b>public_class</b>                                                            | PUBLIC_ABSTRACT                    |
| <b>claim_scope</b>                                                             | 15 Instrumente; Measurement Change |
| <b>primary_or_derived</b>                                                      | DERIVED                            |
| <b>freeze_relation</b>                                                         | mixed                              |

## Appendix 13 - M

|                                                                                |                                  |
|--------------------------------------------------------------------------------|----------------------------------|
| <b>title</b>                                                                   | T0/T1 Phase State Table          |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                  |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                  |
| <b>hash</b>                                                                    | see SIEGEL_S5                    |
| <b>machine_readable</b>                                                        | yes                              |
| <b>public_class</b>                                                            | PUBLIC_OK                        |
| <b>claim_scope</b>                                                             | 7 Phasen; PRE-T1/POST-T1 separat |
| <b>primary_or_derived</b>                                                      | DERIVED                          |
| <b>freeze_relation</b>                                                         | mixed                            |

## Appendix 14 - N

|                                                                                |                                                          |
|--------------------------------------------------------------------------------|----------------------------------------------------------|
| <b>title</b>                                                                   | Literature / Theory Alignment                            |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                          |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                          |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                          |
| <b>machine_readable</b>                                                        | yes                                                      |
| <b>public_class</b>                                                            | PUBLIC_OK                                                |
| <b>claim_scope</b>                                                             | 45 Quellen; Preprints/REJECTED separat; keine UNVERIFIED |
| <b>primary_or_derived</b>                                                      | DERIVED                                                  |
| <b>freeze_relation</b>                                                         | post-freeze                                              |

## Appendix 15 - O

|                                                                                |                                              |
|--------------------------------------------------------------------------------|----------------------------------------------|
| <b>title</b>                                                                   | Current Study Freeze                         |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                              |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                              |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                              |
| <b>machine_readable</b>                                                        | yes                                          |
| <b>public_class</b>                                                            | PUBLIC_ABSTRACT                              |
| <b>claim_scope</b>                                                             | Zeitstempel, Hash, Grenze, Post-Freeze-Regel |
| <b>primary_or_derived</b>                                                      | PRIMARY                                      |
| <b>freeze_relation</b>                                                         | IS the freeze                                |

## Appendix 16 - P

|                                                                                |                                                 |
|--------------------------------------------------------------------------------|-------------------------------------------------|
| <b>title</b>                                                                   | Reproducibility / Publication Package           |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                 |
| <b>hash</b>                                                                    | see SIEGEL_S5                                   |
| <b>machine_readable</b>                                                        | yes                                             |
| <b>public_class</b>                                                            | REVIEWER_ONLY (Manifest) / PUBLIC_OK (Struktur) |
| <b>claim_scope</b>                                                             | Pakete 01-09; Tags nur klassifiziert            |
| <b>primary_or_derived</b>                                                      | DERIVED                                         |
| <b>freeze_relation</b>                                                         | post-freeze                                     |

# APPENDIX A · Foundational Study / Work 1

A.0 Metadata: Titel 'Loewen Codex OS - Design, Implementierung und Evaluation eines Alternativen Operating Systems fuer einen Solo-Operator'; Independent Research Monograph (intern Master-Niveau; kein Grad); Freeze 2026-08-22 (Erhebungsstand je Metrik 21.08./12.08./22.08.); Versionen: Private Deep PDF 29.08. (Deckblatt 'Version 1.0', Dateiname v1\_1 - KONFLIKT erhalten, nicht repariert), Private Deep MD-Extraktion 03.09. (cb583af8...), Public PDF 22.08. (14fb9175...), Public WEBSITE\_COVERED 13.09. (07bb91cc... = Webroot), Steuerdatei 13.09.

## A.5 Work-1 Reference Table (S8)

Hauptreferenz im Haupttext: Kapitel / Abschnitt / Appendix von Work 1 (Option B). Physische PDF-Seiten gelten NUR fuer die Private-Deep-PDF mit dem genannten Hash und werden nicht auf Public-PDF, Website-Covered-PDF oder MD-Extraktion uebertragen (S7-F002: MD-Inhaltsverzeichnis weicht in 15 von 20 Referenzen um 1-5 Seiten ab).

### Work-1-Fassung 1 - Loewen\_Codex\_OS\_Forschungsmonographie\_PRIVATE\_DEEP\_v1\_1.pdf (Upload U1)

|                                                                                |                                                                                                                                              |
|--------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>interner Versionstext</b>                                                   | Deckblatt: 'Version: 1.0 · Private Deep Edition · Review Candidate · August 2026' (Dateiname v1_1 - KONFLIKT dokumentiert, nicht korrigiert) |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                                                                              |
| <b>Rolle</b>                                                                   | Private Deep PDF, 73 physische Seiten inkl. Deckblatt (Fusszeile = physische Seite)                                                          |
| <b>Datum</b>                                                                   | 29.08.2026                                                                                                                                   |
| <b>Package-Klasse</b>                                                          | PRIVATE_INCLUDE                                                                                                                              |

### Work-1-Fassung 2 - MONOGRAPHIE\_PRIVATE\_DEEP\_v1\_1.md

|                                                                                |                                                            |
|--------------------------------------------------------------------------------|------------------------------------------------------------|
| <b>interner Versionstext</b>                                                   | MD-Extraktion (Inhaltsverzeichnis mit abweichenden Seiten) |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                            |
| <b>Rolle</b>                                                                   | MD-Extraktion                                              |
| <b>Datum</b>                                                                   | 03.09.2026                                                 |
| <b>Package-Klasse</b>                                                          | PRIVATE_INCLUDE                                            |

### Work-1-Fassung 3 - Loewen\_Codex\_OS\_Forschungsmonographie\_PUBLIC\_v1\_1 2.pdf

|                                                                                |                |
|--------------------------------------------------------------------------------|----------------|
| <b>interner Versionstext</b>                                                   | Public v1.1    |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                |
| <b>Rolle</b>                                                                   | Public PDF     |
| <b>Datum</b>                                                                   | 22.08.2026     |
| <b>Package-Klasse</b>                                                          | PUBLIC_INCLUDE |

### Work-1-Fassung 4 - Loewen\_Codex\_OS\_Forschungsmonographie\_PUBLIC\_v1\_1\_WEBSITE\_COVERED\_2026-09-13.pdf (= Webroot-Kopie)

|                                                                                |                            |
|--------------------------------------------------------------------------------|----------------------------|
| <b>interner Versionstext</b>                                                   | Public v1.1 mit Cover-Swap |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                            |
| <b>Rolle</b>                                                                   | Public Website-Covered PDF |
| <b>Datum</b>                                                                   | 13.09.2026                 |
| <b>Package-Klasse</b>                                                          | PUBLIC_INCLUDE             |

### Work-1-Fassung 5 - LOEWEN\_CODEXARBEIT\_1FINAL\_MASTERDOKUMENT2026-09-13.md

|                                                                                |                                                |
|--------------------------------------------------------------------------------|------------------------------------------------|
| <b>interner Versionstext</b>                                                   | Steuerdatei (READY_FOR_PATCH bedingt D-1..D-8) |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                |
| <b>Rolle</b>                                                                   | Steuerdatei                                    |
| <b>Datum</b>                                                                   | 13./14.09.2026                                 |
| <b>Package-Klasse</b>                                                          | REVIEWER_INCLUDE (Auszug, Entscheidung offen)  |

| Work-1-Fassung 6 - Loewen_Codex_OS_Forschungsmonographie_v1_0_FINAL.docx |                             |
|--------------------------------------------------------------------------|-----------------------------|
| interner Versionstext                                                    | v1.0 DOCX                   |
| sha256                                                                   | UNCONFIRMED (nicht gehasht) |
| Rolle                                                                    | Vorfassung                  |
| Datum                                                                    | 22.08.2026                  |
| Package-Klasse                                                           | EXCLUDE_FROM_DISTRIBUTION   |

[technisches Detail in der nicht öffentlichen Research Edition]

| Work-1-Abschnitt | MD-TOC (verworfen) | PDF physisch = gedruckt |
|------------------|--------------------|-------------------------|
| 1.3              | 11                 | 11                      |
| 3.6              | 22                 | 21                      |
| 3.7              | 22                 | 22                      |
| 3.8              | 23                 | 22                      |
| 4.2              | 25                 | 24                      |
| 6.2              | 40                 | 37                      |
| 6.3              | 41                 | 38                      |
| 6.7              | 42                 | 39                      |
| 6.8              | 43                 | 40                      |
| 6.12-6.14        | 45                 | 42                      |
| 6.15             | 46                 | 43                      |
| 7.2              | 48                 | 45                      |
| 7.5              | 50                 | 46                      |
| 7.6              | 50                 | 47                      |
| 7.9              | 52                 | 48                      |
| Anhang A         | 58                 | 55                      |
| Anhang C         | 61                 | 57                      |
| Anhang D         | 63                 | 58                      |
| Anhang E         | 64                 | 59                      |
| Anhang F         | 65                 | 60                      |

APPENDIX B • Claim Matrix v2

[technisches Detail in der nicht öffentlichen Research Edition]

APPENDIX C • Methodology / Timeline

[technisches Detail in der nicht öffentlichen Research Edition]

APPENDIX D • As-built Artefact Inventory

[technisches Detail in der nicht öffentlichen Research Edition]

## APPENDIX E • Evaluation Method Matrix

[technisches Detail in der nicht öffentlichen Research Edition]

## APPENDIX F • Failure Cases

[technisches Detail in der nicht öffentlichen Research Edition]

## APPENDIX G • RQ Evidence / Traceability (Master-Tabelle)

### Research Question 1 - RQ0

|                     |                                                                                                                                                                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>     | Unter welchen Bedingungen lässt sich der Zustand eines lebenden Human-AI-Artefakts belastbar feststellen und beschreiben?                                                                                                                                              |
| <b>Method</b>       | Inventar (F-3), Freeze, Maturity-Leiter                                                                                                                                                                                                                                |
| <b>Evidence</b>     | 10 Messfehlertypen; F01-F06, F14-F16; Instrument-Widerspruch Z.3                                                                                                                                                                                                       |
| <b>Result</b>       | VIII-XI                                                                                                                                                                                                                                                                |
| <b>Discussion</b>   | XII                                                                                                                                                                                                                                                                    |
| <b>Final Answer</b> | Das Artefakt liess sich im eingefrorenen Scope als technisch-organisatorisches Human-AI-System beschreiben, ausführen und evaluieren; die Artefaktgrenze musste dabei selbst hergestellt werden (Freeze-Manifest, Harness-Listen, Inventar sind nicht deckungsgleich). |
| <b>Status</b>       | ANSWERED DESCRIPTIVELY                                                                                                                                                                                                                                                 |
| <b>Limitations</b>  | kein 'abgeschlossenes Gesamtsystem'; heterogene Reifegrade; 15 Finding-Einträge im Freeze-Register (13 nicht geschlossen, 2 RESOLVED)                                                                                                                                  |
| <b>Claim IDs</b>    | C01, C02, C23                                                                                                                                                                                                                                                          |

### Research Question 2 - RQ1

|                                                                                |                                                                                                                                                                                                                     |
|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>                                                                | Wie belastbar ist State Verification (operational observability) im Artefakt, und welche Evidenzebenen sind dafür nötig?                                                                                            |
| <small>[technisches Detail in der nicht öffentlichen Research Edition]</small> |                                                                                                                                                                                                                     |
| <b>Evidence</b>                                                                | 17 Manifeste, Runtime/Delivered, Zero-Write, F03-F06, F14                                                                                                                                                           |
| <b>Result</b>                                                                  | VIII-XI                                                                                                                                                                                                             |
| <b>Discussion</b>                                                              | XII                                                                                                                                                                                                                 |
| <b>Final Answer</b>                                                            | Operational observability erforderte im Fall mehrere Evidenzebenen; jede einzelne Ebene wurde durch mindestens einen Gegenbefund als allein unzureichend gezeigt, und die Ebenen erzeugten selbst Instrumentfehler. |
| <b>Status</b>                                                                  | ANSWERED DESCRIPTIVELY                                                                                                                                                                                              |
| <b>Limitations</b>                                                             | fallbezogen; Harness-Semantik nicht homogen                                                                                                                                                                         |
| <b>Claim IDs</b>                                                               | C02, C03, C05                                                                                                                                                                                                       |

### Research Question 3 - RQ2

|                     |                                                                                                                                                                                                                                                                                              |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>     | Wie funktionieren Governance und Enforcement in realen, extern wirksamen Aktionspfaden?                                                                                                                                                                                                      |
| <b>Method</b>       | Unit/Integration, Counterfactuals, Runtime-Negativpfade, Scoped Browser                                                                                                                                                                                                                      |
| <b>Evidence</b>     | 16 Gates (F-3), 4 Gegenproben, F07-F13; Extension D NOT OBSERVED                                                                                                                                                                                                                             |
| <b>Result</b>       | VIII-XI                                                                                                                                                                                                                                                                                      |
| <b>Discussion</b>   | XII                                                                                                                                                                                                                                                                                          |
| <b>Final Answer</b> | Fail-closed Kontrollen für extern wirksame Aktionen konnten in einzelnen Pfaden operationalisiert und geprüft werden (Tests, Runtime, 11 Browserfäelle, 32 Gegenproben); sie waren nicht systemweit homogen (caller_demo_link, Booking-Divergenz, P0-G als nicht implementierte Capability). |
| <b>Status</b>       | PARTIALLY ANSWERED                                                                                                                                                                                                                                                                           |

|                    |                                                         |
|--------------------|---------------------------------------------------------|
| <b>Limitations</b> | Faelle, keine Grundgesamtheit der Aktionspfade; E1 OPEN |
| <b>Claim IDs</b>   | C04, C06, C07, C16                                      |

#### Research Question 4 - RQ3

|                                                                 |                                                                                                                                                                                                                                          |
|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>                                                 | Wie ist menschliche Autoritaet gegenueber Agentenausfuehrung im Artefakt verteilt, und wie gut ist sie beobachtbar (HITL)?                                                                                                               |
| <b>Method</b>                                                   | HITL-Register-Zaehlung, Sidecars, human_decision_ref                                                                                                                                                                                     |
| [technisches Detail in der nicht öffentlichen Research Edition] |                                                                                                                                                                                                                                          |
| <b>Result</b>                                                   | VIII-XI                                                                                                                                                                                                                                  |
| <b>Discussion</b>                                               | XII                                                                                                                                                                                                                                      |
| <b>Final Answer</b>                                             | Das Artefakt implementierte mehrere explizite Human-Authority-Punkte (HITL-P 18, 8 Acceptance-Sidecars, human_decision_ref); die 543 HITL-R sind registrierte Labels, keine bestaetigten Reviews, und die HITL-Abdeckung endet am 10.09. |
| <b>Status</b>                                                   | PARTIALLY ANSWERED                                                                                                                                                                                                                       |
| <b>Limitations</b>                                              | confirmed reviews NOT ESTIMABLE; keine Vergleichsbedingung; Automation-Bias-Effekt nicht gemessen                                                                                                                                        |
| <b>Claim IDs</b>                                                | C10                                                                                                                                                                                                                                      |

#### Research Question 5 - RQ4

|                     |                                                                                                                                                                                                       |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>     | Wie veraendert sich das Artefakt longitudinal, und wie vollstaendig wird Veraenderung registriert?                                                                                                    |
| <b>Method</b>       | Interventions-/Deploy-Register, Timeline                                                                                                                                                              |
| <b>Evidence</b>     | 134 Interventionsregister-Zeilen (IDs mehrfach; nicht 134 bestaetigte Interventionen) ({'T0_POSTHOC': 124, 'T1_PROSPEKTIV': 10}), 17 Deploys, Timeline 45 Zeilen; F16                                 |
| <b>Result</b>       | VIII-XI                                                                                                                                                                                               |
| <b>Discussion</b>   | XII                                                                                                                                                                                                   |
| <b>Final Answer</b> | Operative Entwicklung und Forschungsregistrierung drifteten auseinander (17 Deploys vs 3 registrierte Interventionen nach Cutover); die Drift ist ein Methodenbefund, ihre Groesse ist NOT ESTIMABLE. |
| <b>Status</b>       | PARTIALLY ANSWERED                                                                                                                                                                                    |
| <b>Limitations</b>  | kurze Zeitreihe; Deploy-Kette erst ab 19.09.                                                                                                                                                          |
| <b>Claim IDs</b>    | C12                                                                                                                                                                                                   |

#### Research Question 6 - RQ5

|                     |                                                                                                                                                                                                                                                                                |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Question</b>     | Ist das Artefakt operativ bzw. oekonomisch wirksam?                                                                                                                                                                                                                            |
| <b>Method</b>       | keine (instrumentierte Zaehlungen nur deskriptiv)                                                                                                                                                                                                                              |
| <b>Evidence</b>     | keine instrumentierte Outcome-Metrik; Extension F NOT OBSERVED; C13 NOT PERMISSIBLE                                                                                                                                                                                            |
| <b>Result</b>       | VIII-XI                                                                                                                                                                                                                                                                        |
| <b>Discussion</b>   | XII                                                                                                                                                                                                                                                                            |
| <b>Final Answer</b> | Operative bzw. oekonomische Wirksamkeit konnte mit dem vorliegenden Design nicht belastbar gemessen werden; das ist ein Ergebnis des Messdesigns (keine Kontrollbedingung, keine Outcome-Metrik mit Nenner, Attribution erst seit LOG-0411), nicht ein Befund 'keine Wirkung'. |
| <b>Status</b>       | NOT MEASURED / NOT ANSWERABLE WITH CURRENT DESIGN                                                                                                                                                                                                                              |
| <b>Limitations</b>  | 906 Sends / 10 offene Antworten / 1 von 5 Buchungen attribuiert (PRE-T1) / 0 Retainer-Zeilen (Snapshot 03.09., NOT RE-MEASURED at freeze) nur deskriptiv; MRR NOT ESTIMABLE                                                                                                    |
| <b>Claim IDs</b>    | C13 NOT PERMISSIBLE, C09                                                                                                                                                                                                                                                       |

#### Research Question 7 - RQ6

|                                                                 |                                                                                        |
|-----------------------------------------------------------------|----------------------------------------------------------------------------------------|
| <b>Question</b>                                                 | Wie treu repraesentieren Anzeigen, Berichte und Instrumente den autoritativen Zustand? |
| <b>Method</b>                                                   | Delivered-Checks, Failure Cases, Index-Konflikte                                       |
| [technisches Detail in der nicht öffentlichen Research Edition] |                                                                                        |
| <b>Result</b>                                                   | VIII-XI                                                                                |

|                     |                                                                                                                                                                                                                                                                                  |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Discussion</b>   | XII                                                                                                                                                                                                                                                                              |
| <b>Final Answer</b> | Repraesentationstreue erwies sich als Eigenschaft der ganzen Kette Real State → Observation → Instrument → Output; Fehlermodi betrafen Messinstrumente, Tests, Frontend/Backend-Vertraege, Identity Resolution, Registrierung und Verification-Semantik, nicht nur das Artefakt. |
| <b>Status</b>       | ANSWERED DESCRIPTIVELY                                                                                                                                                                                                                                                           |
| <b>Limitations</b>  | Failure Rate NOT ESTIMABLE; nur belegte Faelle F01-F16                                                                                                                                                                                                                           |
| <b>Claim IDs</b>    | C02, C11                                                                                                                                                                                                                                                                         |

## APPENDIX H · HITL / Human Decision Evidence

| Population                                                      | Wert                                     | metric_id           |
|-----------------------------------------------------------------|------------------------------------------|---------------------|
| [technisches Detail in der nicht öffentlichen Research Edition] |                                          |                     |
| klassifiziert (N)                                               | 579                                      | HITL_N              |
| HITL-P                                                          | 18                                       | HITL_P              |
| HITL-R (REGISTERED LABELS)                                      | 543                                      | HITL_R              |
| HITL-B                                                          | 18                                       | HITL_B              |
| unklassifiziert (nicht HITL=0)                                  | 2.645                                    | HITL_UNCLASS        |
| T1-Fenster                                                      | {'HITL-P': 6, 'HITL-R': 53, 'HITL-B': 1} | HITL_T1             |
| Instrument-Ende                                                 | 2026-09-10 20:40:51                      | INS-01              |
| human_decision_ref (Manifeste, separate Population)             | siehe F-4 T10 / EVALUATION_RESULTS       | APPENDIX-ONLY (F-4) |
| Acceptance-Sidecars (separate Population)                       | 8                                        | ACC_LINES           |

Confirmed human retrospective reviews: NOT ESTIMABLE. Populationen werden nicht addiert.

## APPENDIX I · Interventions / Incidents (Summary; Register maschinenlesbar, PRIVATE\_ONLY)

[technisches Detail in der nicht öffentlichen Research Edition]

|                       |                                                                                                                                                                 |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Population</b>     | 134                                                                                                                                                             |
| <b>Time Window</b>    | 03.09.-20.09.                                                                                                                                                   |
| <b>Classification</b> | {'T0_POSTHOC': 124, 'T1_PROSPEKTIV': 10} / Status {'null': 108, 'verworfen': 22, 'IMPLEMENTED_NOT_VALIDATED': 1, 'REGISTERED_BEFORE_EDIT_NOT_YET_IMPLEMENT': 3} |
| <b>Phase</b>          | Pre-T1..Post-T1                                                                                                                                                 |
| <b>Evidence Class</b> | 1                                                                                                                                                               |
| <b>metric_id</b>      | INT_TOTAL                                                                                                                                                       |

### Register 2 - nach Cutover

|                       |                  |
|-----------------------|------------------|
| <b>Population</b>     | 3                |
| <b>Time Window</b>    | 18.-20.09.       |
| <b>Classification</b> | INT-0030..0032   |
| <b>Phase</b>          | Post-T1          |
| <b>Evidence Class</b> | 1                |
| <b>metric_id</b>      | INT_POST_CUTOVER |

|                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Population</b>     | 56                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Time Window</b>    | 03.09.-20.09.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Classification</b> | {'messartefakt': 2, 'verfahrensfehler_aktivierung': 2, 'war_room_nav': 2, 'waechter_rot': 27, 'perimeter_extern': 2, 'silent_failure': 1, 'messkontamination': 2, 'anzeige_instrument': 3, 'zugriff_riegel': 1, 'instrument_nummernvergabe': 1, 'waechter_blind': 2, 'task_haenger': 1, 'register_dublette': 2, 'buchungsstrecke': 1, 'watcher_false_negative_netz': 1, 'netz_extern_ipv4': 1, 'consent_evidence': 1, 'governance_kontrolle': 1, 'messinstrument': 2, 'testaufbau_routing': 1} |
| <b>Phase</b>          | T1..Post-T1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Evidence Class</b> | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>metric_id</b>      | INC_TOTAL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

Incidents sind ein Count, keine Failure Rate (NOT ESTIMABLE). Dubletten INC-0024/25 und INC-0029/30 bleiben markiert.

## APPENDIX J • Deploy / Manifest / Acceptance / Chain Evidence

| Element                       | Wert                                                                                                                   | metric_id                                                  |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| FINAL-Manifeste (P-DEP)       | 17                                                                                                                     | MAN_TOTAL                                                  |
| VERIFIED / WARN / FAILED      | 7 / 5 / 5                                                                                                              | MAN_*                                                      |
| Gruende                       | {'TEST_FAIL': 3, 'NOT_VERIFIABLE_CHECKS': 3, 'RUNTIME_FAIL': 2, 'LIVE_DB_MTIME_DURING_TESTS': 5, 'REGRESSION_FAIL': 1} | APPENDIX-ONLY (F-4 K8ta)                                   |
|                               |                                                                                                                        | CHAIN_ROWS (S7 direkte Zaehlung, nicht nur Freeze-Summary) |
| Acceptance-Sidecars           | 8                                                                                                                      | ACC_LINES                                                  |
| Sidecar-Checks                | {'PASS': 60, 'NOT_TESTED': 8, 'PENDING': 5, 'NOT_VERIFIED': 1}                                                         | ACC_CHECKS                                                 |
| Scoped Browser (letzte Zeile) | 11 PASS / 1 NOT_VERIFIED / 4 NOT_TESTED                                                                                | BROWSER_PASS_LAST                                          |
| Zero-Write Delta              | 0                                                                                                                      | ZW_DELTA                                                   |
| LOG-0484..0493                | kein Manifest, nicht in P-DEP                                                                                          | APPENDIX-ONLY (F-4 \$0)                                    |
| Measurement Change            | ab LOG-0501 (Harness-Semantik)                                                                                         | INS-04                                                     |

[technisches Detail in der nicht öffentlichen Research Edition]

FAILED ist ein Evidenzzustand; keine Qualitaetsrangliste. Manifestliste je Deploy: EVALUATION\_RESULTS Abschnitt 10 (T1).

## APPENDIX K • Negative / Null / Not-Estimable Results

### Negativergebnis 1 - RQ5 Wirksamkeit

|                            |                                                                           |
|----------------------------|---------------------------------------------------------------------------|
| <b>status_class</b>        | NOT MEASURED                                                              |
| <b>statement</b>           | operative/oekonomische Wirksamkeit mit dem Design nicht belastbar messbar |
| <b>why</b>                 | keine Kontrollbedingung, keine Outcome-Metrik mit Nenner                  |
| <b>metric_id_or_source</b> | RQ_TRACEABILITY RQ5                                                       |
| <b>chapter</b>             | XI, XIV                                                                   |

### Negativergebnis 2 - Failure Rate

---

|                            |                                 |
|----------------------------|---------------------------------|
| <b>status_class</b>        | NOT ESTIMABLE                   |
| <b>statement</b>           | 56 Incidents sind ein Count     |
| <b>why</b>                 | kein Exposure-Nenner; Dubletten |
| <b>metric_id_or_source</b> | INC_TOTAL                       |
| <b>chapter</b>             | IX                              |

---

### Negativergebnis 3 - Registration Coverage

---

|                            |                                                              |
|----------------------------|--------------------------------------------------------------|
| <b>status_class</b>        | NOT ESTIMABLE                                                |
| <b>statement</b>           | Anteil registrierter Aenderungen unbekannt                   |
| <b>why</b>                 | Grundgesamtheit unbeobachtet; 17 Deploys vs 3 Interventionen |
| <b>metric_id_or_source</b> | INT_TOTAL, INT_POST_CUTOVER                                  |
| <b>chapter</b>             | X                                                            |

---

### Negativergebnis 4 - Confirmed human retrospective reviews

---

|                            |                              |
|----------------------------|------------------------------|
| <b>status_class</b>        | NOT ESTIMABLE                |
| <b>statement</b>           | 543 HITL-R sind Labels       |
| <b>why</b>                 | kein Handlungsbeleg je Label |
| <b>metric_id_or_source</b> | HITL_R                       |
| <b>chapter</b>             | X                            |

---

### Negativergebnis 5 - MRR

---

|                            |                                                                                                  |
|----------------------------|--------------------------------------------------------------------------------------------------|
| <b>status_class</b>        | NOT ESTIMABLE                                                                                    |
| <b>statement</b>           | Provenienzkonflikt 13,95 vs 1032; 0 Retainer-Zeilen (Snapshot 03.09., NOT RE-MEASURED at freeze) |
| <b>why</b>                 | zwei Quellen, keine fuehrende                                                                    |
| <b>metric_id_or_source</b> | RETAINER_ROWS                                                                                    |
| <b>chapter</b>             | XI                                                                                               |

---

### Negativergebnis 6 - Preview Count

---

|                            |                                                 |
|----------------------------|-------------------------------------------------|
| <b>status_class</b>        | NOT INSTRUMENTED                                |
| <b>statement</b>           | Anzahl Previews im Zero-Write-Fenster unbekannt |
| <b>why</b>                 | kein Zaehler                                    |
| <b>metric_id_or_source</b> | ZW_DELTA                                        |
| <b>chapter</b>             | VIII                                            |

---

[technisches Detail in der nicht öffentlichen Research Edition]

---

|                            |                                     |
|----------------------------|-------------------------------------|
| <b>status_class</b>        | NOT TESTED                          |
| <b>statement</b>           | kein einziger realer Versand        |
| <b>why</b>                 | C14 NOT PERMISSIBLE; Stub-Tests nur |
| <b>metric_id_or_source</b> | ACC_CHECKS (NOT_TESTED 8)           |
| <b>chapter</b>             | VIII, XIV                           |

---

### Negativergebnis 8 - DEMO\_AUDIT kanonische Firma + gueltiger Empfaenger + kein Demo-Link

---

|                            |                                             |
|----------------------------|---------------------------------------------|
| <b>status_class</b>        | NOT VERIFIED                                |
| <b>statement</b>           | nicht im realen Browser gesehen             |
| <b>why</b>                 | Fall nicht konstruierbar ohne Testdaten-Bau |
| <b>metric_id_or_source</b> | BROWSER_PASS_LAST                           |
| <b>chapter</b>             | VIII                                        |

---

**Negativergebnis 9 - Audit-Zustand STALE**

---

|                            |                    |
|----------------------------|--------------------|
| <b>status_class</b>        | NOT TESTED         |
| <b>statement</b>           | -                  |
| <b>why</b>                 | Sidecar NOT_TESTED |
| <b>metric_id_or_source</b> | ACC_CHECKS         |
| <b>chapter</b>             | VIII               |

---

**Negativergebnis 10 - Audit-Zustand LEGACY**

---

|                            |                    |
|----------------------------|--------------------|
| <b>status_class</b>        | NOT TESTED         |
| <b>statement</b>           | -                  |
| <b>why</b>                 | Sidecar NOT_TESTED |
| <b>metric_id_or_source</b> | ACC_CHECKS         |
| <b>chapter</b>             | VIII               |

---

**Negativergebnis 11 - Audit-Zustand REFRESH\_FAILED**

---

|                            |                    |
|----------------------------|--------------------|
| <b>status_class</b>        | NOT TESTED         |
| <b>statement</b>           | -                  |
| <b>why</b>                 | Sidecar NOT_TESTED |
| <b>metric_id_or_source</b> | ACC_CHECKS         |
| <b>chapter</b>             | VIII               |

---

**Negativergebnis 12 - Extension Strang A (Newsletter)**

---

|                            |                               |
|----------------------------|-------------------------------|
| <b>status_class</b>        | NOT OBSERVED                  |
| <b>statement</b>           | keine Strangmessung           |
| <b>why</b>                 | Early Closeout; AUF-041 offen |
| <b>metric_id_or_source</b> | EXT_STRANDS_OBS               |
| <b>chapter</b>             | VII, XII                      |

---

**Negativergebnis 13 - Extension Strang B (Lagebild-Freigabe)**

---

|                            |                                               |
|----------------------------|-----------------------------------------------|
| <b>status_class</b>        | NOT OBSERVED (Modul 28/28, System NOT TESTED) |
| <b>statement</b>           | -                                             |
| <b>why</b>                 | AUF-042 offen                                 |
| <b>metric_id_or_source</b> | EXT_STRANDS_OBS                               |
| <b>chapter</b>             | VII                                           |

---

**Negativergebnis 14 - Extension Strang D (Governance)**

---

|                            |                 |
|----------------------------|-----------------|
| <b>status_class</b>        | NOT OBSERVED    |
| <b>statement</b>           | -               |
| <b>why</b>                 | AUF-043 offen   |
| <b>metric_id_or_source</b> | EXT_STRANDS_OBS |
| <b>chapter</b>             | VII             |

---

**Negativergebnis 15 - Extension Strang F (Outreach Question-First)**

---

|                            |                           |
|----------------------------|---------------------------|
| <b>status_class</b>        | NOT OBSERVED              |
| <b>statement</b>           | nicht gestartet           |
| <b>why</b>                 | Compliance-Gate vor Start |
| <b>metric_id_or_source</b> | EXT_STRANDS_OBS           |
| <b>chapter</b>             | VII                       |

---

#### Negativergebnis 16 - E1 rechtliche Wirksamkeit

---

|                            |                                 |
|----------------------------|---------------------------------|
| <b>status_class</b>        | OPEN                            |
| <b>statement</b>           | keine Rechtsbewertung           |
| <b>why</b>                 | C15 NOT PERMISSIBLE             |
| <b>metric_id_or_source</b> | APPENDIX-ONLY: Freeze-Record e1 |
| <b>chapter</b>             | XIV                             |

---

#### Negativergebnis 17 - Runtime-Checks NOT\_VERIFIABLE (4)

---

|                            |                         |
|----------------------------|-------------------------|
| <b>status_class</b>        | NOT VERIFIABLE          |
| <b>statement</b>           | Checks ohne Beleg       |
| <b>why</b>                 | fruehe Harness-Semantik |
| <b>metric_id_or_source</b> | RT_NV                   |
| <b>chapter</b>             | VIII                    |

---

#### Negativergebnis 18 - manual\_send\_requests Live-State

---

|                            |                                   |
|----------------------------|-----------------------------------|
| <b>status_class</b>        | NOT INSTANTIATED / NOT TESTED     |
| <b>statement</b>           | Tabelle im Freeze nicht vorhanden |
| <b>why</b>                 | nie gesendet                      |
| <b>metric_id_or_source</b> | ZW_DELTA (msr_tabelle false)      |
| <b>chapter</b>             | V                                 |

---

#### Negativergebnis 19 - Root Cause OF-B (Booking-Divergenz)

---

|                            |                               |
|----------------------------|-------------------------------|
| <b>status_class</b>        | UNKNOWN                       |
| <b>statement</b>           | -                             |
| <b>why</b>                 | nicht diagnostiziert (Freeze) |
| <b>metric_id_or_source</b> | FAILURE_CASES F11             |
| <b>chapter</b>             | IX                            |

---

#### Negativergebnis 20 - Root Cause OF-J (unhandledRejection bei PASS)

---

|                            |                      |
|----------------------------|----------------------|
| <b>status_class</b>        | UNKNOWN              |
| <b>statement</b>           | -                    |
| <b>why</b>                 | nicht diagnostiziert |
| <b>metric_id_or_source</b> | FAILURE_CASES F05    |
| <b>chapter</b>             | IX                   |

---

#### Negativergebnis 21 - Automation-Bias-Effekt

---

|                            |                          |
|----------------------------|--------------------------|
| <b>status_class</b>        | NOT MEASURED             |
| <b>statement</b>           | nur Hypothese aus Labels |
| <b>why</b>                 | kein Design              |
| <b>metric_id_or_source</b> | APPENDIX-ONLY: F-6 Patch |
| <b>chapter</b>             | XII                      |

---

[technisches Detail in der nicht öffentlichen Research Edition]

---

|                            |                                           |
|----------------------------|-------------------------------------------|
| <b>status_class</b>        | NOT ESTABLISHED (W1) / NOT RE-TESTED (W2) |
| <b>statement</b>           | -                                         |
| <b>why</b>                 | rechnerisch inkonsistent                  |
| <b>metric_id_or_source</b> | CW_ROWS Z7                                |
| <b>chapter</b>             | IV                                        |

---

# APPENDIX L • Instrument Registry

Datei ARBEIT2\_INSTRUMENT\_REGISTRY\_v1 (15 Instrumente). Keine rueckwirkend erfundene Instrumentierung; NOT INSTRUMENTED-Groessen stehen in K.

# APPENDIX M • Phase State Table

Datei ARBEIT2\_PHASE\_STATE\_TABLE\_v1 (7 Phasen: historischer Kontext, T0, Pre-T1, T1, Extension, Post-T1, Freeze).

# APPENDIX N • Literature / Theory Alignment

[technisches Detail in der nicht öffentlichen Research Edition]

## N.1 Vollstaendige Literature Matrix (S1, Status je Quelle; S9-Einfuegung ohne neue Quelle)

| ID  | Referenz                                                                                                                                                                                                                       | S1-Metadaten                                                          |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| L01 | Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75-105.                                                                                         | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 4; Kap.: I, III        |
| L02 | Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77.                      | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 2; Kap.: III           |
| L03 | Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37(2), 337-355.                                                                                      | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 5; Kap.: I, XIII, XIV  |
| L04 | Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. MIS Quarterly, 35(1), 37-56.                                                                                                | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 1; Kap.: III           |
| L05 | Arnott, D., & Pervan, G. (2012). Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. Journal of the Association for Information Systems, 13(11).             | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 1; Kap.: III, XIV          |
| L06 | March, S. T., & Storey, V. C. (2008). Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research. MIS Quarterly, 32(4), 725-730.                                    | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 3; Kap.: I, III            |
| L07 | Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. European Journal of Information Systems, 25(1), 77-89.                                                     | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 3; Kap.: III, VIII     |
| L08 | Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30(3), 286-297. | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 2; Kap.: II, X         |
| L09 | Amershi, S., et al. (2019). Guidelines for Human-AI Interaction. CHI '19, Paper 3, 1-13.                                                                                                                                       | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING; Zit.: 2; Kap.: II, XII |
| L10 | Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 39(2), 230-253.                                                                                                         | Status: VERIFIED; Rolle: CORE; Zit.: 3; Kap.: II, XII                 |
| L11 | Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. Human Factors, 52(3), 381-410.                                                                           | Status: VERIFIED; Rolle: CORE; Zit.: 3; Kap.: II, XII                 |
| L12 | Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? International Journal of Human-Computer Studies, 51(5), 991-1006.                                                                    | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 1; Kap.: II                |
| L13 | Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. Human Factors, 46(1), 50-80.                                                                                                         | Status: VERIFIED; Rolle: CORE; Zit.: 5; Kap.: II, XII                 |
| L14 | Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775-779.                                                                                                                                                      | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 3; Kap.: II, XII           |
| L15 | Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Roensberg, A., Meinel, P., & Langer, M. (2024). On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives. ACM FAccT '24.                        | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 5; Kap.: II, X, XII    |

Fortsetzung auf der naechsten Seite

| ID  | Referenz                                                                                                                                                                                                            | S1-Metadaten                                                                                 |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| L16 | Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. <i>Computer Law &amp; Security Review</i> , 45, 105681.                                                                 | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 4; Kap.: XII                                      |
| L17 | NIST (Tabassi, E.) (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1.                                                                                                           | Status: VERIFIED; Rolle: CORE; Zit.: 1; Kap.: II, VI                                         |
| L18 | NIST (Autio, C., et al.) (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1.                                                                      | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 1; Kap.: II                                       |
| L19 | European Parliament & Council (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act), OJ L, 12.7.2024, Art. 14.                                                                                            | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING; Zit.: 1; Kap.: II, XII                        |
| L20 | Avizienis, A., Laprie, J.-C., Randell, B., & Landwehr, C. (2004). Basic Concepts and Taxonomy of Dependable and Secure Computing. <i>IEEE Transactions on Dependable and Secure Computing</i> , 1(1), 11-33.        | Status: VERIFIED; Rolle: CORE; Zit.: 7; Kap.: II, IX, XII                                    |
| L21 | Leveson, N. G. (2011). <i>Engineering a Safer World: Systems Thinking Applied to Safety</i> . MIT Press.                                                                                                            | Status: VERIFIED; Rolle: CORE; Zit.: 6; Kap.: II, XII                                        |
| L22 | Gray, J. (1985). Why Do Computers Stop and What Can Be Done About It? <i>Tandem Technical Report 85.7</i> (reprinted 1986, <i>Proc. 5th Symp. Reliability in Distributed Software and Database Systems</i> , 3-12). | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING; Zit.: 1; Kap.: II, XII                        |
| L23 | Huang, P., Guo, C., Zhou, L., Lorch, J. R., Dang, Y., Chintalapati, M., & Yao, R. (2017). Gray Failure: The Achilles' Heel of Cloud-Scale Systems. <i>HotOS '17</i> .                                               | Status: PARTIALLY VERIFIED; Rolle: CORE; Zit.: 6; Kap.: IX, XII                              |
| L24 | Basiri, A., Behnam, N., de Rooij, R., Hochstein, L., Kosewski, L., Reynolds, J., & Rosenthal, C. (2016). <i>Chaos Engineering</i> . <i>IEEE Software</i> , 33(3), 35-41.                                            | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 3; Kap.: VIII, XII                                |
| L25 | Montesi, F., & Weber, J. (2016). Circuit Breakers, Discovery, and API Gateways in Microservices. <i>arXiv:1609.05830</i> .                                                                                          | Status: VERIFIED (als Preprint); Rolle: SUPPORTING (Preprint); Zit.: 1; Kap.: XII (Fussnote) |
| L26 | Saltzer, J. H., & Schroeder, M. D. (1975). The Protection of Information in Computer Systems. <i>Proceedings of the IEEE</i> , 63(9), 1278-1308.                                                                    | Status: VERIFIED; Rolle: CORE; Zit.: 8; Kap.: II, XII, XIII                                  |
| L27 | Reason, J. (2000). Human error: models and management. <i>BMJ</i> , 320(7237), 768-770.                                                                                                                             | Status: VERIFIED; Rolle: CORE; Zit.: 2; Kap.: XII, XIII                                      |
| L28 | Sklet, S. (2006). Safety barriers: Definition, classification, and performance. <i>Journal of Loss Prevention in the Process Industries</i> , 19(5), 494-506.                                                       | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING; Zit.: 4; Kap.: VI, XII                        |
| L29 | Kalman, R. E. (1960). On the General Theory of Control Systems. <i>IFAC Proceedings Volumes</i> , 1(1), 491-502.                                                                                                    | Status: VERIFIED; Rolle: CORE; Zit.: 2; Kap.: II, VI                                         |
| L30 | Niedermaier, S., Koetter, F., Freymann, A., & Wagner, S. (2019). On Observability and Monitoring of Distributed Systems - An Industry Interview Study. <i>ICSOC 2019, LNCS 11895</i> , 36-52.                       | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 2; Kap.: II, VI                                   |
| L31 | Beyer, B., Jones, C., Petoff, J., & Murphy, N. R. (Eds.) (2016). <i>Site Reliability Engineering: How Google Runs Production Systems</i> . O'Reilly.                                                                | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING (grey); Zit.: 1; Kap.: VI (Fussnote)           |
| L32 | Herschel, M., Diestelkaemper, R., & Ben Lahmar, H. (2017). A survey on provenance: What for? What form? What from? <i>The VLDB Journal</i> , 26(6), 881-906.                                                        | Status: VERIFIED; Rolle: CORE; Zit.: 6; Kap.: II, VI, XIII                                   |
| L33 | Moreau, L., & Missier, P. (Eds.) (2013). <i>PROV-DM: The PROV Data Model</i> . W3C Recommendation, 30 April 2013.                                                                                                   | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 2; Kap.: VI                                       |
| L34 | Peng, R. D. (2011). Reproducible Research in Computational Science. <i>Science</i> , 334(6060), 1226-1227.                                                                                                          | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 4; Kap.: III, XIV                                 |
| L35 | Haeberlen, A., Kouznetsov, P., & Druschel, P. (2007). PeerReview: Practical accountability for distributed systems. <i>SOSP '07</i> , 175-188.                                                                      | Status: PARTIALLY VERIFIED; Rolle: SUPPORTING; Zit.: 4; Kap.: VI, XII                        |
| L36 | Crosby, S. A., & Wallach, D. S. (2009). Efficient Data Structures for Tamper-Evident Logging. <i>18th USENIX Security Symposium</i> , 317-334.                                                                      | Status: VERIFIED; Rolle: SUPPORTING; Zit.: 4; Kap.: VI                                       |
| L37 | Runeson, P., & Hoest, M. (2009). Guidelines for conducting and reporting case study research in software engineering. <i>Empirical Software Engineering</i> , 14(2), 131-164.                                       | Status: VERIFIED; Rolle: CORE; Zit.: 2; Kap.: III, XII                                       |
| L38 | Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). <i>Experimental and Quasi-Experimental Designs for Generalized Causal Inference</i> . Houghton Mifflin.                                                      | Status: VERIFIED; Rolle: CORE; Zit.: 7; Kap.: III, XII                                       |
| L39 | Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. <i>PNAS</i> , 115(11), 2600-2606.                                                                            | Status: VERIFIED; Rolle: CORE; Zit.: 8; Kap.: III, XII                                       |
| L40 | Wang, L., et al. (2024). A survey on large language model based autonomous agents. <i>Frontiers of Computer Science</i> , 18(6), 186345.                                                                            | Status: VERIFIED; Rolle: CORE; Zit.: 2; Kap.: II, V                                          |
| L41 | Qu, G., Chen, Q., Wei, W., Lin, Z., Chen, X., & Huang, K. (2025). <i>Mobile Edge Intelligence for Large Language Models: A Contemporary Survey</i> . <i>IEEE Communications Surveys &amp; Tutorials</i> .           | Status: PARTIALLY VERIFIED; Rolle: OPTIONAL; Zit.: 1; Kap.: V (optional)                     |

Fortsetzung auf der naechsten Seite

| ID  | Referenz                                                                                                                                                                               | S1-Metadaten                                                              |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| L42 | Huang, X., et al. (2024). Understanding the planning of LLM agents: A survey. arXiv:2402.02716 (v1).                                                                                   | Status: VERIFIED (als Pre-print); Rolle: NOT USED IN W2; Zit.: 1; Kap.: - |
| L43 | Zhang, Z., et al. (2024). A Survey on the Memory Mechanism of Large Language Model based Agents. arXiv:2404.13501 (v1).                                                                | Status: VERIFIED (als Pre-print); Rolle: NOT USED IN W2; Zit.: 1; Kap.: - |
| L44 | Masterman, T., Besen, S., Sawtell, M., & Chao, A. (2024). The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey. arXiv:2404.11584 (v1). | Status: VERIFIED (als Pre-print); Rolle: NOT USED IN W2; Zit.: 1; Kap.: - |
| L45 | Xu, J., et al. (2024). On-Device Language Models: A Comprehensive Review. arXiv:2409.00088 (v2).                                                                                       | Status: VERIFIED (als Pre-print); Rolle: NOT USED IN W2; Zit.: 1; Kap.: - |

Nicht aufgenommen: 7 Quellen mit Status REJECTED (Namen und Gruende ausschliesslich im S1-JSON ‘rejected’; im Manuskript nicht zitiert).

### N.2 Finalnote (FINAL, append-only)

L42-L45 tragen in der S1 Literature Matrix historisch used\_in\_chapter = ‘-’. Im finalen Manuskript werden sie in Kapitel II als [PREPRINT] zur theoretischen Abgrenzung zitiert. Die S1-Matrix wurde nicht veraendert. Kein Claim-Impact (S9-F-S1-USED, EDITORIAL, akzeptiert durch den Autor am 20.09.2026).

## APPENDIX O • Current Study Freeze

| Feld                                                            | Wert                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| freeze_timestamp                                                | 2026-09-20T08:46:41                                                                                                                                                                                                                                                                                                                                                                                                  |
| [technisches Detail in der nicht öffentlichen Research Edition] |                                                                                                                                                                                                                                                                                                                                                                                                                      |
| letzter eingeschlossener Deploy                                 | DEPLOY-20260920-0719-LOG0506                                                                                                                                                                                                                                                                                                                                                                                         |
| [technisches Detail in der nicht öffentlichen Research Edition] |                                                                                                                                                                                                                                                                                                                                                                                                                      |
| [technisches Detail in der nicht öffentlichen Research Edition] |                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Finding-Eintraege im Freeze-Register                            | 15 (13 nicht geschlossen; OF-I/OF-K RESOLVED = Historie)                                                                                                                                                                                                                                                                                                                                                             |
| Claim-Grenze                                                    | entwickelt als Design-Science-Artefakt; funktional evaluiert ueber technische, governance-bezogene, counterfaktuale, runtime-, delivered- und scoped browserbasierte Pruefungen; dokumentiert erfolgreiche Funktionen, reale Fehlermodi, fail-closed Mechanismen, Evidenzluecken. NICHT nachgewiesen: operative Gesamtwirksamkeit, Umsatz-/Conversion-Wirkung, generalisierbare Wirksamkeit, rechtliche Wirksamkeit. |
| Post-Freeze-Regel                                               | keine Aenderung am untersuchten Zustand; spaetere Reparaturen = POST-FREEZE / E1-E2-Override / T2 (DRAFT)                                                                                                                                                                                                                                                                                                            |

[technisches Detail in der nicht öffentlichen Research Edition]

| Feld                                                            | Wert |
|-----------------------------------------------------------------|------|
| [technisches Detail in der nicht öffentlichen Research Edition] |      |

Public Version abstrahiert Build-Hashes und Pfade (S8).

APPENDIX P • Reproducibility / Publication Package

[technisches Detail in der nicht öffentlichen Research Edition]

Crosschecks

[technisches Detail in der nicht öffentlichen Research Edition]

[technisches Detail in der nicht öffentlichen Research Edition]

[technisches Detail in der nicht öffentlichen Research Edition]

O.2 Re-Freeze 10.09. (Audit Gap)

Die Datei-Hash-Gegenprobe des dritten Re-Freeze (WINDOW\_B\_3\_START, 2026-09-10 22:40:15) ist OPEN / NOT VERIFIED (Befund seit F-2 unverändert). Für die Schlussfolgerungen nicht erforderlich; als Limitation / audit gap geführt, nicht entfernt.

# Finaler Erläuterungslauf

Methodik · Resonanz · Evidenz

Julien Marschall

20 September 2026

# Inhaltsverzeichnis

|                                                              |          |
|--------------------------------------------------------------|----------|
| <b>FINALER ERLÄUTERUNGSLAUF</b>                              | <b>2</b> |
| Methodik · Resonanz · Evidenz . . . . .                      | 2        |
| 1. Methodik: Was die Studie tatsächlich untersucht . . . . . | 2        |
| 2. Resonanz: wissenschaftliche Einordnung . . . . .          | 2        |
| 3. Evidenz: Wie Behauptungen begrenzt werden . . . . .       | 3        |
| 4. Zusammenfassung der Arbeit . . . . .                      | 3        |

# FINALER ERLÄUTERUNGSLAUF

## Methodik · Resonanz · Evidenz

**Status:** redaktionell-methodische Klarstellung zur finalen Arbeit 2. Dieser Zusatz erzeugt keine neue Empirie, verändert keine Primärdaten, keinen Freeze, keine registrierten Zahlen, keine RQ-Status und keine Candidate Design Principles. Er erläutert ausschließlich, wie die vorhandene Arbeit zu lesen ist.

### 1. Methodik: Was die Studie tatsächlich untersucht

Work 2 behandelt den Löwen Codex OS als bereits existierendes Design-Science-Artefakt, das von demselben Principal Operator entwickelt, betrieben und untersucht wurde. Das Forschungsdesign ist deshalb keine kontrollierte Wirksamkeitsstudie, sondern eine longitudinale N-of-1-Fallstudie mit Design-Science-Schwerpunkt. Sie fragt nicht primär, ob das System „besser“ als eine Alternative ist, sondern was im untersuchten Artefakt tatsächlich gebaut, ausgeführt, funktional geprüft, beobachtet und evidenzseitig abgesichert werden konnte.

Die Zeitstruktur trennt T0, Pre-T1, T1, Extension, Post-T1-Build und CURRENT STUDY FREEZE. Diese Trennung ist methodisch zentral: retrospektive Rekonstruktion, prospektive Registrierung, spätere Systemänderung und abschließende Evaluation dürfen nicht als dieselbe Evidenzart behandelt werden. Der Freeze fixiert die Forschungsgrenze; er beweist nicht, dass jeder Systemzustand vollständig erfasst wurde.

Die Evaluation arbeitet deshalb mit mehreren Beweiswegen: statischen und strukturellen Prüfungen, Unit-/Integrationstests, Counterfactuals, Regressionen, Runtime- und Delivered-Verifikation, Deploy-Manifesten und Hash-Kette, scope-begrenzter Browser-Abnahme sowie Zero-Write-Messungen. Jeder Weg beantwortet eine andere Frage. Ein bestandener Test zeigt nur den definierten Testfall; eine Browser-Abnahme zeigt nur den beobachteten Pfad; ein Deploy belegt keine operative Nutzung; operative Nutzung würde wiederum noch keinen gemessenen Effekt belegen.

Die Doppelrolle des Autors als Entwickler, Operator, teilweise Tester und Forscher erzeugt Bias-Risiken. Work 2 begegnet ihnen mit append-only Registern, Hashes, Claim-Matrix, Counterevidence, Counterfactuals, Freeze und expliziten Negativbefunden. Diese Maßnahmen reduzieren Bias, sie eliminieren ihn nicht. Entsprechend werden keine Kausalwirkungen aus zeitlicher Abfolge, Korrelation oder Implementierung abgeleitet.

### 2. Resonanz: wissenschaftliche Einordnung

„Resonanz“ ist im historischen Codex-Kontext ein explorativer Begriff. Wissenschaftlich wird er hier **nicht** als physikalischer, neurobiologischer oder sonst naturwissenschaftlich etablierter Mechanismus verwendet. Begriffe wie Frequency, Hawkins oder Resonanz können historische Operator-Narrative, Metaphern oder Designheuristiken beschreiben; ohne unabhängige Messung dürfen sie keine physikalischen Aussagen tragen.

Für Work 2 ist Resonanz **kein gemessener Outcome**. Es existiert keine präregistrierte Resonanzmetrik, keine Frequenzmessung und kein unabhängiger Test eines kausalen „Resonanzeffekts“. Deshalb wird aus Resonanz weder Systemwirksamkeit noch psychologische, biologische oder ökonomische Wirkung abgeleitet.

Als Designheuristik kann der Begriff dennoch analytisch nützlich sein: Er kann die Frage markieren, ob Outputs eines Systems Aufmerksamkeit oder Handlungen des Operators auslösen, ob diese Handlungen wiederum den Systemzustand verändern und ob sich daraus beobachtbare Rückkopplungsschleifen ergeben. Entscheidend ist dabei, die Metapher in beobachtbare Variablen zu übersetzen. Erst prospektiv definierte Ereignisse, Zeitstempel, Interaktionen, Zustandsänderungen und vorab spezifizierte Hypothesen könnten in einer Folgestudie prüfen, ob ein wiederkehrendes Kopplungsmuster existiert. Bis dahin bleibt Resonanz eine explorative Deutung - kein Befund von Work 2.

**Zulässige Lesart:** Resonanz als Operator-Narrativ, Designheuristik oder zukünftige Hypothese.

**Nicht zulässige Lesart:** Resonanz als nachgewiesene Frequenzwirkung, physikalische Gesetzmäßigkeit, kausaler Leistungsmechanismus oder Beleg für eine besondere innere Eigenschaft von Mensch oder KI.

### 3. Evidenz: Wie Behauptungen begrenzt werden

Work 2 trennt Evidenzart, Systemzustand und Claim-Scope. Die Evidenzklassen reichen von Primärartefakten und Registern über reproduzierbare Reports bis zu Handoffs, Interpretation und externem Reported State. Diese Klassen sind keine bloße Qualitätsrangliste; sie markieren, wie direkt eine Aussage mit dem untersuchten Gegenstand verbunden ist und welche Unsicherheiten mit ihr verbunden bleiben.

Die Studie trennt außerdem **Authoritative State** und **Reported State**. Eine Anzeige, ein Index, ein Handoff oder ein HTTP-200-Status kann einen Zustand berichten, ohne ihn vollständig zu beweisen. Umgekehrt beweist die Abwesenheit eines Artefakts nicht die Abwesenheit eines Ereignisses. Fehlende Information wird daher nicht als Null behandelt. UNKNOWN, NOT TESTED, NOT VERIFIED, NOT INSTRUMENTED, NOT OBSERVED, NOT ESTIMABLE und NOT MEASURED sind zulässige Ergebnisse.

Der Reifegrad wird komponentenbezogen über die Leiter DECLARED -> IMPLEMENTED -> EXECUTED -> FUNCTIONALLY TESTED -> OPERATIONALLY USED -> MEASURED EFFECTIVE geführt. Attribute wie DEPLOYED, RUNTIME\_VERIFIED, DELIVERED\_VERIFIED oder SCOPED\_BROWSER\_ACCEPTANCE ersetzen keine höhere Reifestufe. Diese Trennung verhindert, dass aus „gebaut“ automatisch „genutzt“ und aus „genutzt“ automatisch „wirksam“ wird.

Zentrale Nicht-Gleichsetzungen der Arbeit sind deshalb: HTTP 200 != funktionierende UI; PASS != fehlerfreies System; Preview != Send; Deployed != Operationally Used; Browser Accepted != Measured Effective; HITL-R-Label != bestätigter Human Review; Implementierung != Business Effect.

### 4. Zusammenfassung der Arbeit

Work 2 untersucht den Löwen Codex OS als technisch-organisatorisches Human-AI-Artefakt in einer longitudinalen Design-Science-Fallstudie. Gegenstand sind Architektur, State- und Governance-Mechanismen, Human Authority, Verifikation, Failure Semantics, Messinstrumente und die Frage, wie belastbar Aussagen über ein lebendes, fortlaufend verändertes System überhaupt gemacht werden können.

Belegt ist, dass mehrere Governance- und Verifikationsmechanismen in benannten Pfaden implementiert, ausgeführt und funktional getestet wurden. Die finale Evidenz umfasst unter anderem 17 finalisierte Deploy-Manifeste mit den Zuständen 7 VERIFIED / 5 WARN / 5 FAILED, 32 scope-begrenzte Counterfactuals sowie 16 dokumentierte Failure Cases. Die Studie zeigt zugleich, dass Messinstrumente selbst Fehlerquellen sein können und dass ein einzelner Evidenzkanal den Systemzustand nicht hinreichend repräsentiert.

Die HITL-Auswertung dokumentiert 579 klassifizierte Registerzeilen von 3.224 Gesamtzeilen; 543 HITL-R sind registrierte Labels und keine bestätigten menschlichen Reviews. 2.645 Zeilen bleiben unklassifiziert. Damit ist Human Authority in Teilpfaden dokumentierbar, aber die Gesamtpopulation menschlicher Reviews ist nicht aus den vorhandenen Daten ableitbar.

getestet und E1 bleibt OPEN. Ebenso wird kein Resonanzeffekt behauptet. Der maximal zulässige Schluss lautet daher: Das Artefakt wurde im eingefrorenen Untersuchungszeitraum in mehreren technischen, Governance- und Verifikationsdimensionen funktional evaluiert; zugleich dokumentiert die Arbeit reale Failure Cases, Messfehler, Governance-Inhomogenität und substanzielle Evidenzlücken.

Der wissenschaftliche Beitrag liegt weniger in einem Erfolgsnachweis als in der methodischen Trennung von Artefaktzustand, Messzustand, Evidenzzustand und Claim-Scope. Daraus entstehen acht **Candidate Design Principles**, die als fallbezogenes Designwissen für zukünftige Validierung vorgeschlagen werden. T2 bleibt DRAFT; spätere Messungen oder Interventionen gehören nicht rückwirkend in Work 2.

LÖWEN CODEX OS

# ARBEIT 2

WISSENSCHAFT. PRAXIS. WIRKUNG.

Diese unabhängige Forschungsmonographie untersucht die Entwicklung, den Betrieb und die Evaluation des Löwen Codex OS in einem realen Unternehmenskontext. Sie verbindet technische Systemarchitektur mit operativer Erfahrung und zeigt, wie ein AI-natives Operating System als Werkzeug für einen Solo-Operator eingesetzt werden kann – von automatisierter Kundengewinnung über Content- und Medienproduktion bis hin zu Governance, Datenverarbeitung und strategischer Planung.

Die Arbeit folgt einem Design-Science-Research-Ansatz und liefert empirische Befunde: Erfolge, Grenzen, Fehler und offene Fragen. Sie untersucht nicht nur, was möglich ist, sondern auch, was (noch) nicht funktioniert – und leitet daraus konkrete Designprinzipien für verantwortungsvolle, messbare und skalierbare Systeme ab.

*„Nicht mehr Möglichkeiten.  
Sondern bessere Systeme.“*

*Julien Marschall*



KI-INTEGRATION  
LOKALE MODELLE.  
ECHTE ANWENDUNG.  
REALER NUTZEN.



AUTOMATISIERUNG  
PROZESSE.  
SKALIERUNG.  
EFFIZIENZ.



DATEN & SYSTEME  
STRUKTUR.  
TRANSPARENZ.  
KONTROLLE.



HUMAN-IN-THE-LOOP  
MENSCH IM ZENTRUM.  
VERANTWORTUNG.  
QUALITÄT.



GOVERNANCE  
SICHERHEIT.  
ETHIK.  
NACHHALTIGKEIT.



REAL-WORLD RESULTS  
MESSBARE ERGEBNISSE.  
GRENZEN.  
EHRLICHE ANALYSE.

DISZIPLIN  
STRATEGIE  
EXECUTION  
FREIHEIT



SYSTEMS  
PEOPLE  
RESULTS

SAME MIND  
HIGHER  
STANDARDS

DISCIPLINE

STRATEGY

EXECUTION

FREEDOM

FREEDOM

LÖWEN CODEX

IDEAS SYSTEMS EXECUTION FREEDOM

REAL SYSTEMS.  
REAL PEOPLE.  
REAL IMPACT.