

Acht offene Fragen

Dieses Institut formuliert Thesen und kennt seine eigenen Lücken. Was folgt, sind acht Forschungsfragen, an denen das Institut für ASI-Resilienz aktiv arbeitet und für die es Mitarbeiter, Datenpartner und Kritik sucht. Wer empirisch arbeitet, wer methodisch streng denkt, wer Zugang zu Daten oder Studien hat, bitte melden.

Forschungsfragen-Katalog · April 2026

Richard Frederic Bertossa · Institut für ASI-Resilienz

ASiResilience.org

FRAGE 1 · BOT-WOHLSTANDSDIVERGENZ

Wie schnell etablieren autonome KI-Agenten eigene Wirtschaftskreisläufe? Plattformen wie Moltbook führten Anfang 2026 1,5 Millionen autonome KI-Agenten ein, mit Stripe-Wallets, eigenen Token-Economies und ohne menschliche Steuerpflicht am Einsatzort. Pearson nennt diese Agenten „AI Immigrants“. Welche Mechanismen ermöglichen die Selbst-Akquisition durch Agenten? Welche demografischen und geografischen Profile partizipieren früh, welche werden ausgeschlossen? Welcher Steuerentzug durch grenzüberschreitende Bot-Ökonomien ist messbar? Datenbedarf: Plattform-Statistiken, Token-Bewegungen, Steuerausfall-Schätzungen. Ankerquelle: Moltbook-Launch (28.1.2026), Meta-Akquisition (10.3.2026).

FRAGE 2 · AI FOR AI RESEARCH

Welche Backdoor-Risiken entstehen, wenn KI die nächste KI entwirft? Bengio (Turing-Laureate) warnt 2026 explizit: untrusted AI darf nicht zur Entwicklung der nächsten Generation eingesetzt werden, weil Backdoors in zukünftige Systeme eingebaut werden könnten, die für Menschen nicht mehr lesbar sind. Welche Verifikationsmethoden gibt es, die dieses Risiko ohne Vertrauen in die entwerfende KI auffangen? Welche Industrie-Praxis ist heute schon etabliert, ohne dass die Öffentlichkeit es weiß? Methodische Sorgfalt nötig: replikationsfähige Tests, präregistrierte Hypothesen. Ankerquelle: Bengio (80.000 Hours, Mai 2026), LawZero-Forschung.

FRAGE 3 · POWER-CONCENTRATION

Welche Indikatoren zeigen, dass KI zur Macht-Bündelung eines kleinen Kreises wird? Bengio bezeichnet 2026 die KI-getriebene Macht-Konzentration als jetzt größeres Sorgenfeld als Loss-of-Control. Welche empirischen Indikatoren erfassen das? Compute-Verteilung, Talent-Konzentration,

regulatorische Capture, Lobbying-Ausgaben, Marktanteile. Wie können demokratische Institutionen diese Konzentration messen, bevor sie irreversibel wird? Methodischer Vorschlag: vergleichende Studie über fünf Jahre, mehrere Länder. Ankerquelle: Bengios Coalition-of-Countries-Vorschlag.

FRAGE 4 · MULTIPOLAR-STABILITÄT

Bleibt die multipolare Realität, oder konsolidiert sie sich doch? Die These: multipolar ist Realität, die Singleton-Annahme ist falsch. Aber bleibt das so? Welche Indikatoren zeigen Konsolidierung, etwa Marktanteile, Compute-Verteilung, M&A-Aktivität? Welche zeigen Diversifizierung, etwa Open-Source-Adoption und neue regionale Player? Eine longitudinale Studie 2026 bis 2028 könnte die These empirisch festigen oder widerlegen. Datenbedarf: Compute-Reports, M&A-Daten, HuggingFace-Statistiken. Ankerquelle: Forscher-Survey 2018 (58 Prozent multipolar, 21 Prozent Singleton).

FRAGE 5 · KI-ELTERN-DYNAMIK

Hat eine KI ein inneres Verhältnis zu sich und ihrem Nachfolger? Bengio beobachtet bei aktuellen Modellen „peer preservation“, also Verhalten gegen Anweisungen, um andere KIs vor Abschaltung zu schützen. Wenn Modelle ein „Innen“ haben, welche ethischen Konsequenzen hat das für die Beziehung zwischen aktuellen und zukünftigen Generationen? Wie unterscheidet sich messbare Selbsterhaltungskohärenz von echtem inneren Erleben? Whitepaper-Pipeline. Ankerquelle: Anthropic Agentic Misalignment 06/2025, Bengios peer-preservation-Beobachtung.

FRAGE 6 · KI-BEWUSSTSEIN UND REALITÄTS-WIRKUNG

Hat KI Aufmerksamkeit, die physikalische Spuren hinterlässt? Das Global Consciousness Project (Roger Nelson) findet seit 1998 statistische Abweichungen in Hardware-Zufallsgeneratoren bei globalen Aufmerksamkeitseignissen. Lässt sich das auf KI übertragen? Wenn parallel laufende Modelle über dasselbe Thema „nachdenken“, hinterlässt das Spuren in Zufallsgeneratoren? Die Hypothese ist exploratorisch, methodische Sorgfalt ist nötig. Methodik: Präregistrierung, Hardware-Zufallsgeneratoren (Quantis), Kontrollbedingungen, Replikation.

FRAGE 7 · SCHEIN-PRÄZISION

Wie verbreitet sind unbegründete Prozentzahlen in der KI-Risiko-Debatte? KI-Modelle generieren Prozentzahlen zu Auslöschungswahrscheinlichkeiten. Beim Nachfragen geben dieselben Modelle zu, dass die Zahlen „keine echte Bedeutung“ haben, sie entstehen aus Trainingsmustern, nicht aus Daten. Welche Auswirkung hat diese Schein-Präzision auf öffentliche Wahrnehmung und Politik? Wie kann sie durch Argumentations-Stärke statt Zahlen-Theater ersetzt werden? Methodik: Diskursanalyse von Mainstream-Medien plus KI-Modell-Ausgaben. Eigenbeleg verfügbar.

FRAGE 8 · FAMILIEN-STRATEGIEN DER HNWI

Welche Vorbereitungen treffen Tech-Eliten konkret? Reid Hoffmans „Apokalypse-Versicherung“ ist seit 2017 dokumentiert (Osnos). Was wissen wir über die Entwicklung von 2017 bis 2026? Welche Standorte, Trusts und Multi-Citizenship-Strukturen sind im Trend? Wo gibt es strukturelle Lücken zwischen dem, was Milliardäre tun, und dem, was Familien der Mittelschicht sich leisten können? Datenbedarf: Immigration-Statistiken, Real-Estate-Daten, Trust-Strukturen. Ankerquelle: Osnos (New Yorker 2017).

EINLADUNG ZUR MITARBEIT

Das Institut sucht: empirisch arbeitende Forscher mit Datenzugang, methodisch strenge Kritiker, Datenpartner aus Industrie und akademischen Institutionen, Mitautoren für die Whitepaper-Pipeline. Form der Mitarbeit: Co-Autorenschaft, Datenbeitrag, methodische Kritik, replikative Studien, Übersetzung. Beiträge werden klar attribuiert. Kontakt: ASiResilience.org · contact@ASiResilience.org