Comparing Notes on Trustworthy AI 2025

Part 2: AI Agents in Practice: Risk vs. Reward - Is It Worth It?

German version below

Host: appliedAI Institute for Europe
Date: 1 July 2025
Location: House of Communication, Munich

The second event built on the foundations of Session 1 to dive deeper into the practical realities of AI agent implementation. The central question: given the potential benefits and risks, when and how should organisations implement AI agents? Concrete use cases, governance frameworks, and legal realities were examined together.

The panel

Dr. Antonia von Appen is a Senior Associate in Digital Business at Noerr. She provided the legal and compliance lens, emphasising the need for governance frameworks before deployment and the importance of clear liability allocation in contracts.

Jutta Juliane Meier is Founder and CEO of Identity Valley. She brought the digital responsibility perspective, advocating for measurable frameworks for trust and highlighting the importance of addressing responsible AI during early stages of the ML lifecycle.

Albert Pujol is Founder and CEO of AI Shepherds. He presented a concrete use case in Quality Management Systems where AI finds similar problems, suggests containment actions, and cuts time to completion by around 50%.

Marc Alexander Kuhn is Investment Manager at UVC Partners. He stressed the importance of avoiding vendor lock-in, maintaining ownership of data trails, and evaluating AI startups by time-to-value, automation degree, and data uniqueness.

The panel was moderated by **Claudia Baumgartner**, Trustworthy AI Expert at the appliedAI Institute for Europe.

Discussion

Do not race - have a genuine use case

The panel opened with a clear message from Antonia von Appen: organisations should not implement AI agents simply because they are trending. Legal compliance requires patience and thoroughness, and genuine business needs must drive the decision - not competitive pressure.

Marc Kuhn observed that the competitive landscape will change rapidly and AI innovation is accelerating faster than many organisations can absorb. The investors

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.

perspective: tools that generate value quickly without weeks of integration, high automation degrees, and proprietary data are the real differentiators.

Risk, liability, and the AI Act

Dr. von Appen outlined the legal liability framework. Risk depends on the environment: internal applications carry lower risk than customer-facing ones, and sector-specific risks apply in medical and financial contexts. Governance must span three levels: strategic, operational, and structural (including role definitions).

The panel assessed the AI Act's readiness for agents: while the Act does not mention AI agents specifically, it will apply - an Agent is an AI System. Critically, in multi-agent systems, if one agent is high-risk, all other relevant agents are also considered high-risk.

Key challenges: employment law, liability, GDPR, and copyright all apply simultaneously. The AI Act has led to confusion among lawyers; as a world-first regulation, there will be gaps, and future legal requirements are hard to predict.

Treating AI agents as digital employees

Marc Kuhn provocatively framed AI agents as being like employees - and argued they should be treated as such. Companies can automate 80-90% of human work, which is genuinely disruptive. Antonia von Appen agreed: establish appropriate supervision levels, start with clear definitions of what the agent is supposed to do, and consider recourse options with developers. Human oversight becomes even more important with agents than with traditional AI systems.

Albert responded from the startup perspective: he emphasised staying agile without spending heavily on lawyers. His recommendation: start with human-in-the-loop approaches and gradually reduce oversight as confidence grows.

Practical governance and data sovereignty

Albert shared a live example from the morning: the orchestrator did not work and agents responded who were not meant to respond. This illustrated that proper prompting is one of the most effective ways to ensure correct functioning, and that human oversight remains a critical safety mechanism.

On data sovereignty: many companies are moving away from US and China solutions while waiting for good EU alternatives. GDPR aims to keep data in Europe. Jutta Meier emphasised the need to communicate with policymakers to ensure sovereignty through proper policy design.

Getting started: practical steps

Albert's approach: educate yourself, learn proper prompting techniques, participate in an implementation project, then develop your own use case. Antonia's legal framework: before giving an AI agent power, create a governance framework and oversight structures; establish solid contracts between value chain actors. Large models do not

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.

specify certain purposes - they claim they can do everything but accept liability for nothing. Organisations must handle liability and performance issues carefully.

Setting guardrails requires a structured approach: define the agent's role clearly, provide background context, specify tasks precisely, and include examples of how to carry out tasks.

Cost-benefit reality check

Organisations should evaluate AI agents across four primary dimensions: ROI (deployment ease, security controls, human fall-back instances, compliance risks); positive business case (consequences of false decisions, autonomy-based risk assessment); liability considerations (risk acceptance thresholds, multi-agent implications); and impact measurement (value quantification, digital trust, measurability of outcomes).

Key costs often underestimated: people, time, quality investments, total cost of ownership, maintenance requirements, and hallucination management. Time-to-value calculations for employee productivity gains must account for all of these.

Key takeaways

Implementing trustworthy AI agents requires balancing innovation with responsibility. The most successful approach combines:

- Clear problem definition: find your problem first.
- Realistic assessment of both capabilities and limitations.
- Iterative development with continuous learning.
- Appropriate human oversight and prompting scaled to risk levels.
- Cross-functional collaboration between technical, business, and legal teams from the start.
- Contractual clarity on liability and performance expectations.
- Start simple: begin with document processing or internal applications.
- Think strategically: avoid vendor lock-in and maintain data sovereignty.
- Implement governance early: do not wait until problems arise.

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.

Comparing Notes on Trustworthy AI 2025

Teil 2: KI-Agenten in der Praxis: Risiko vs. Reward - Lohnt es sich?

English version above

Gastgeber: appliedAI Institute for Europe
Datum: 1. Juli 2025
Ort: Haus der Kommunikation, München

Das zweite Event tauchte tiefer in die praktischen Realitäten der Implementierung von KI-Agenten ein. Die zentrale Frage: Wann und wie sollten Unternehmen angesichts der potenziellen Vorteile und Risiken KI-Agenten implementieren?

Das Panel

Dr. Antonia von Appen ist Senior Associate im Digital Business bei Noerr. Sie lieferte die rechtliche und Compliance-Perspektive, betonte Governance-Frameworks vor dem Einsatz und die Bedeutung klarer Haftungsverteilung in Verträgen.

Jutta Juliane Meier ist Gründerin und CEO von Identity Valley. Sie brachte die Perspektive der digitalen Verantwortung ein und sprach sich für messbare Vertrauensrahmen aus.

Albert Pujol ist Gründer und CEO von AI Shepherds. Er präsentierte einen konkreten Anwendungsfall in Qualitätsmanagementsystemen: KI findet ähnliche Probleme, schlägt Massnahmen vor und reduziert die Bearbeitungszeit um rund 50%.

Marc Alexander Kühn ist Investment Manager bei UVC Partners. Er betonte die Bedeutung, Anbieter-Lock-in zu vermeiden und Eigentümer der eigenen Datenpfade zu bleiben.

Das Panel wurde moderiert von **Claudia Baumgartner**, Trustworthy AI Expert bei der appliedAI Institute for Europe gGmbH.

Diskussion

Kein Wettrennen - einen echten Anwendungsfall haben

Das Panel eröffnete mit einer klaren Botschaft von Antonia von Appen: Organisationen sollten keine KI-Agenten implementieren, nur weil sie im Trend liegen. Die Einhaltung rechtlicher Vorschriften erfordert Geduld und Gründlichkeit. Echte Geschäftsbedürfnisse müssen die Entscheidung antreiben.

Marc Kühn beobachtete, dass sich das Wettbewerbsumfeld schnell ändern wird und KI-Innovationen schneller zunehmen, als viele Organisationen absorbieren können. Die Investorenperspektive: Tools, die schnell Wert generieren, ohne wochenlange Integration, hohe Automatisierungsgrade und proprietäre Daten, sind die wahren Unterscheidungsmerkmale.

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.

Risiko, Haftung und der AI Act

Dr. von Appen skizzierte den rechtlichen Rahmen: Interne Anwendungen tragen geringeres Risiko als kundenorientierte. Die Governance muss drei Ebenen umfassen: strategisch, operativ und strukturell. Der KI-Act wird auf Agenten angewandt - ein Agent ist ein KI-System. In Multi-Agenten-Systemen gilt: wenn ein Agent als hochriskant eingestuft wird, gelten alle anderen relevanten Agenten ebenfalls als hochriskant.

KI-Agenten als digitale Mitarbeiter behandeln

Marc Kühn formulierte provokativ: KI-Agenten sind wie Mitarbeiter und sollten auch so behandelt werden. Unternehmen können 80-90% der menschlichen Arbeit automatisieren, was wirklich disruptiv sein kann. Antonia von Appen stimmte zu: angemessene Aufsichtsebenen etablieren, mit klaren Definitionen beginnen und Regressmöglichkeiten mit Entwicklern berücksichtigen.

Praktische Governance und Datenhoheit

Albert teilte ein Beispiel: der Orchestrator funktionierte nicht, und Agenten antworteten, die nicht antworten sollten. Dies verdeutlichte: richtiges Prompting ist eine der effektivsten Massnahmen und menschliche Aufsicht bleibt kritisch. Viele Unternehmen distanzieren sich von US-Lösungen und warten auf gute EU-Alternativen. DSGVO zielt darauf ab, Daten in Europa zu halten.

Erste Schritte: Praktische Ratschläge

Alberts Ansatz: Weiterbilden, Prompting-Techniken erlernen, an einem Implementierungsprojekt teilnehmen, eigenen Anwendungsfall entwickeln. Antonias rechtlicher Rahmen: vor dem Einsatz einen Governance-Rahmen und Aufsichtsstrukturen erstellen; solide Verträge zwischen Akteuren der Wertschöpfungskette etablieren. Grosse Modelle übernehmen keine Haftung - Organisationen müssen dies sorgfältig handhaben.

Kosten-Nutzen-Check

Organisationen sollten KI-Agenten anhand von vier Hauptdimensionen bewerten: ROI (Einsatzfähigkeit, Sicherheitskontrollen, menschliche Fallback-Instanzen, Compliance-Risiken); positiver Business Case (Konsequenzen von Fehlentscheidungen, Autonomie-basierte Risikobewertung); Haftungsfragen (Risikoakzeptanzschwellen, Implikationen für Multi-Agenten-Systeme); und Folgenmessung (Wertquantifizierung, digitales Vertrauen, Messbarkeit der Ergebnisse).

Oft unterschätzte Schlüsselkosten: Personal, Zeit, Qualitätsinvestitionen, Gesamtbetriebskosten, Wartungsanforderungen und Halluzinationsmanagement. Bei der Berechnung der Time-to-Value für Produktivitätssteigerungen der Mitarbeiter müssen all diese Faktoren berücksichtigt werden.

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.

Wichtigste Erkenntnisse

Die Implementierung vertrauenswürdiger KI-Agenten erfordert ein Gleichgewicht zwischen Innovation und Verantwortung:

- Klare Problemdefinition - finden Sie zuerst Ihr Problem.
- Realistische Bewertung von Fähigkeiten und Einschränkungen.
- Iterative Entwicklung mit kontinuierlichem Lernen.
- Angemessene menschliche Aufsicht, skaliert auf Risikostufen.
- Funktionsübergreifende Zusammenarbeit von Anfang an.
- Vertragliche Klarheit bezüglich Haftung und Leistungserwartungen.
- Einfach starten: mit Dokumentenverarbeitung oder internen Anwendungen beginnen.
- Governance von Beginn an implementieren: nicht warten, bis Probleme auftreten.

Note: This document includes AI-generated summaries and translations created with the assistance of Gemini. / Dieses Dokument enthält KI-generierte Zusammenfassungen und Übersetzungen, die mit Unterstützung von Gemini erstellt wurden.