Comparing Notes on Trustworthy AI 2025

Part 5: From Theory to Practice: Lessons Learned and Real-World Applications of Trustworthy AI Agents

German version below

Host: appliedAI Institute for Europe
Date: 13 November 2025
Location: House of Communication, Munich

The fifth and final event reflected on the year's learnings and bridged expert theory with real-world enterprise AI implementation. The central conclusion was that responsible AI agent implementation requires balancing innovation with governance, technical capability with human oversight, and efficiency gains with the protection of human agency and the value of work. The format was a fireside chat with 40 participants.

The panel

Claudia Baumgartner (appliedAI Institute) moderated and provided the practical and sociotechnical perspective, synthesizing the series' key lessons.

Dr. Benjamin Lange (LMU/MCML) presented ethics and research perspectives, noting that AI agents amplify existing grey areas and magnify ethical trade-offs.

Bernhard Waltl (Liquid Legal Institute) covered legal and governance dimensions, emphasising that AI must be a tool in the toolbox, not a replacement for humans.

Fernanda Sauca (TUM Think Tank) addressed AI agents and workplace transformation, highlighting that trust must be intentionally designed, not treated as a by-product.

Maximilian Furtmair (aicx) provided the startup implementation perspective, sharing concrete enterprise deployments and five principles for success.

Discussion

Lessons from the 2025 Series

Four key lessons were synthesized: first, identify the problem before deploying the AI agent; second, agents require significant oversight as they lack human judgment and accountability; third, good instructions are necessary but not sufficient for trustworthy agents; and fourth, getting implementation right means building both trust and accountability. Results should be measured beyond mere efficiency to include quality, fairness, and alignment with organizational values.

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.

Startup Implementation

Maximilian Furtmair's core philosophy is to "teach AI how humans work—not teach humans how to work with AI" to drive adoption. Successful enterprise implementation relies on five principles: solving actual use-cases, demonstrating clear measurable value, ensuring seamless integration, maintaining compliance and IT security, and enabling teams without overwhelming them.

Bridging Theory and Practice

The common thread across all five events was the word "trust," evolving from trusting facts to trusting actions. Key tensions included the need for cultural proximity between agent providers and users, and preserving worker agency so people can maintain control and overrule AI decisions. The real skill needed for humans is "calibrated trust"—knowing when to ask the AI—rather than baseline education.

Protecting Human Value

The panel stressed the intentional need to preserve human agency, sense of control, and sense of purpose that comes from work. The concept of "calibrated trust"—appropriate trust levels developed through experience and ongoing adjustment—was identified as crucial.

Key takeaways

- Collaborative, interdisciplinary implementation with proper training and clear rules from the start.
- Balancing technical, business, sociotechnical, and legal perspectives throughout.
- Maintaining meaningful human oversight always.
- Starting small but thinking big about impact.
- Protecting human agency and the value of work.

Core message: The goal is sustainable AI that enhances rather than replaces human capability, requiring a focus on orchestration rather than full automation. Success demands ongoing learning, adaptation, and community collaboration.

Comparing Notes on Trustworthy AI 2025

Teil 5: Von der Theorie zur Praxis - Gelernte Lektionen und reale Anwendungen vertrauenswürdiger KI-Agenten

English version above

Gastgeber: appliedAI Institute for Europe**
Datum: 13. November 2025**
Ort: Haus der Kommunikation, München

Die fünfte und letzte Veranstaltung schloss die Reihe 2025 mit einer Reflexion der Jahreslernkurve und der Überbrückung von Expertentheorie mit realen Unternehmens-KI-Implementierungserfahrungen ab. Die zentrale Erkenntnis war, dass die verantwortungsvolle Implementierung von KI-Agenten ein Gleichgewicht erfordert: zwischen Innovation und angemessener Governance, technischer Leistungsfähigkeit und menschlicher Aufsicht sowie Effizienzsteigerungen und dem Schutz der menschlichen Handlungsfähigkeit und des Arbeitsplatzwertes. Das Format war ein Fireside Chat mit 40 Teilnehmern.

Das Panel

Claudia Baumgartner (appliedAI Institute) moderierte und brachte die praktische und soziotechnische Perspektive ein. Sie synthetisierte die wichtigsten Lektionen der Reihe.

Dr. Benjamin Lange (LMU/MCML) lieferte Ethik- und Forschungsperspektiven und betonte, dass KI-Agenten bestehende Grauzonen verstärken.

Bernhard Waltl (Liquid Legal Institute) behandelte die rechtlichen und Governance-Dimensionen und forderte, dass KI ein Werkzeug im Werkzeugkasten sein muss – kein Ersatz für Menschen.

Fernanda Sauca (TUM Think Tank) adressierte KI-Agenten und Arbeitsplatztransformation und unterstrich, dass Vertrauen bewusst gestaltet werden muss, nicht als Nebenprodukt behandelt werden darf.

Maximilian Furtmair (aicx) lieferte die Start-up-Implementierungsperspektive mit konkreten Unternehmens-KI-Einsätzen.

Diskussion

Lektionen aus der Reihe 2025

Vier Lektionen wurden synthetisiert: Erstens, finden Sie zuerst Ihr Problem, dann bauen oder kaufen Sie den KI-Agenten; Zweitens, KI-Agenten benötigen die gleiche Führung wie Mitarbeiter – und noch mehr Aufsicht; Drittens, gute Anweisungen sind notwendig, aber

nicht ausreichend für vertrauenswürdige Agenten; Viertens, der richtige Weg bedeutet, Vertrauen und Rechenschaftspflicht aufzubauen. Ergebnisse sollten über die bloße Effizienz hinaus gemessen werden, um Qualität, Fairness und Übereinstimmung mit Organisationswerten zu bewerten.

Start-up-Implementierung

Maximilian Furtmairs Idee ist es, der KI beizubringen, wie Menschen arbeiten – nicht den Menschen beizubringen, wie sie mit der KI arbeiten, um eine höhere Akzeptanzrate zu erzielen. Eine erfolgreiche Unternehmensimplementierung basiert auf fünf Prinzipien: Anwendungsfall statt Werkzeug (Lösung tatsächlicher Geschäftsprobleme); klarer Mehrwert; nahtlose Integration in die bestehende Arbeitsumgebung; Einhaltung von Compliance und IT-Sicherheit; und Teams befähigen ohne zu überfordern.

Theorie und Praxis verbinden

Das zentrale Wort über alle fünf Events hinweg war "Vertrauen", das sich vom Vertrauen in Fakten zum Vertrauen in Handlungen entwickelte. Kulturelle Nähe zwischen Agentenanbietern und Nutzern ist entscheidend, und die menschliche Handlungsfähigkeit muss gewahrt bleiben, damit Menschen die Kontrolle behalten und KI-Entscheidungen außer Kraft setzen können. Die eigentliche Fähigkeit, die der Mensch benötigt, ist das "kalibrierte Vertrauen" – zu wissen, wann man die KI fragen muss.

Menschlichen Wert schützen

Die Diskussion betonte die bewusste Notwendigkeit, die menschliche Handlungsfähigkeit, das Gefühl der Kontrolle und den Sinn für den Zweck, der aus der Arbeit entsteht, zu erhalten. Das Konzept des "kalibrierten Vertrauens" – ein angemessenes, durch Erfahrung entwickeltes und kontinuierlich angepasstes Vertrauensniveau – wurde als entscheidend identifiziert.

Wichtigste Erkenntnisse

- Kollaborative, interdisziplinäre Implementierung mit angemessener Schulung und klaren Regeln von Anfang an.
- Abwägung technischer, geschäftlicher, soziotechnischer und rechtlicher Perspektiven durchgehend.
- Stets Aufrechterhaltung einer sinnvollen menschlichen Aufsicht.
- Klein anfangen, aber gross über die Auswirkungen nachdenken.
- Schutz der menschlichen Handlungsfähigkeit und des Wertes der Arbeit.

Kernbotschaft: Das Ziel ist eine nachhaltige KI, die die menschliche Leistungsfähigkeit verbessert, anstatt sie zu ersetzen. Dies erfordert einen Fokus auf Orchestrierung statt vollständiger Automatisierung. Erfolg setzt kontinuierliches Lernen, Anpassung und Zusammenarbeit in der Gemeinschaft voraus.