Bericht zum Workshop „Trustworthy AI and its successful integration into processes“ im Rahmen der CODE-Jahrestagung 2025

Am 09. Juli 2025 fand während der Jahrestagung des Forschungsinstituts CODE ein Workshop zum Thema *Trustworthy AI and its successful integration into processes* statt. Der folgende Bericht soll einen kurzen Überblick über den Workshop und seine Ergebnisse geben.

Ziel des Workshops war ein offener Austausch mit interaktiven Formaten zum Thema. Außerdem gab es hochrangige Impulse aus unterschiedlichen Disziplinen und unterschiedlichen Bereichen der Praxis und des Militärs. Eine ganzheitliche Betrachtung des Themenkomplexes stand im Vordergrund. Im Zuge des Workshops wurden folgende Ergebnisse herausgearbeitet.

Die Verzahnung von Top-Technologien (AI, genAI, agentic AI, weitere) und Menschzentrierung sowie die systematische Gestaltung von menschzentrierten Prozessen und Dienstleistungen ist essenziell!

Vertrauen in KI bei militärischer Entscheidungsunterstützung:
Wenn KI-Systeme Kommandierenden bei Entscheidungsprozessen assistieren, sind diese selten in der Lage, die KI-Leistung selbst zu evaluieren – meist aus Zeitmangel. Daraus ergibt sich die zentrale Frage, wie Nutzer Vertrauen in Systeme entwickeln können, die sie nicht eigenständig validiert haben.

Kontextabhängigkeit und Entwicklungsdynamik von KI-Systemen:
Die Leistungsfähigkeit von KI bleibt stark kontextabhängig. Mit dem Aufkommen großer Sprachmodelle (LLMs) sind deutliche Fortschritte erkennbar – doch offen bleibt die Frage, wohin diese Entwicklung führt und wie sich die Leistungsfähigkeit künftig in komplexen Einsatzumgebungen zeigt.

Ressourcenintensität und Kosten-Nutzen-Abwägung:
KI-Modelle profitieren von langfristigen, umfangreichen Trainingsdaten. Daraus ergeben sich hohe technische und infrastrukturelle Anforderungen (Trainingsressourcen, On-Site-Support). Es gilt, den Zielkonflikt zwischen Effizienzgewinnen und Ressourcenbedarf sorgfältig auszubalancieren.

Philosophisch-ethische Dimension von Vertrauenswürdigkeit:
Zentrale Prinzipien wie Verantwortung, Erklärbarkeit und Transparenz müssen ethisch reflektiert und praktisch umgesetzt werden. Governance-Fragen sind

branchenübergreifend herausfordernd. Orientierung bieten komplexe Regulierungsrahmen aus anderen Domänen, die zeigen, wie eine verantwortungsvolle, risikobewusste KI-Integration gestaltet werden kann.

Militärische Perspektive: Spannungsfeld zwischen Innovation und Risikoaversion:
Der Vergleich zwischen der Innovationsgeschwindigkeit in der Ukraine und der systematischen, risikominimierenden Vorgehensweise der Bundeswehr verdeutlicht den Zielkonflikt zwischen Agilität und Kontrolle.

Das Prinzip des „Human-in-the-Loop“ bleibt essenziell, erfordert jedoch eine klare Definition von „meaningful oversight“ in zunehmend automatisierten Umgebungen.

Cybersecurity als kollektive Verantwortung:
Cybersicherheit ist kein isoliertes Unternehmensziel, sondern ein Gemeinschaftsanliegen. Eine resiliente digitale Infrastruktur erfordert motivierte Akteur:innen, gemeinschaftliche Verantwortung und den Aufbau kollektiver Kompetenz.

Ziel ist ein sicherheitsbewusstes Ökosystem, in dem alle Beteiligten zum Schutz beitragen.

Die Prognose für die weitere Entwicklung der Technologie ist insofern, dass Herausforderung und Potential zugleich der Technologie dual-use-Anwendungen sind, die den Anforderungen und Bedürfnissen unterschiedlicher Stakeholder gerecht werden, dabei gleichzeitig auch das große Ganze und gesellschaftliche Relevante oftmals erstmalig ganzheitlich betrachten und in positive Gestaltung transferieren.

Echte Menschzentrierung ist bei fast gar keiner Technologie wirklich im Fokus. Das gilt es gerade im dual-use-Kontext aufzugreifen. Gemeinsam mit Wiss./Praxis/Militär und auch interdisziplinär mit Blick über den nationalen Tellerrand.