

VERITAS

VERtrauenswürdige Identitätsfeststellung durch Transparente Agenten-Systeme

Programm / Ausschreibung	KIRAS, Kooperative F&E-Projekte, KIRAS CS Kooperative F&E-Projekte (CS KFE_2025)	Status	laufend
Projektstart	01.10.2026	Projektende	30.09.2028
Zeitraum	2026 - 2028	Projektlaufzeit	24 Monate
Barwert der Projektförderung	€ 678.278		
Keywords	KI-Agenten;Copilot;E-Governance;Explainable AI		

Projektbeschreibung

VERITAS (Vertrauenswürdige Identitätsfeststellung durch Transparente Agenten-Systeme)

Ausgangssituation, Problematik und Motivation

Das Asylwesen ist durch komplexe, zeitkritische und datenintensive Verfahren geprägt. SachbearbeiterInnen des Bundesamts für Fremdenwesen und Asyl (BFA) bewerten täglich eine Flut unstrukturierter Informationen (Identitätsdokumente, mehrstündige Einvernahmeprotokolle, Länderberichte, Altakten). Identitäts-, Alters- und Herkunftstäuschungen sind dabei enorme Herausforderungen. Bislang erfolgen quellenübergreifende Konsistenzprüfungen meist manuell. Angesichts deutlich wachsender Datenmengen und strenger gesetzlicher Fristen (max. 6 Monate) ist dies kognitiv extrem belastend und fehleranfällig. Da das Personal nicht im gleichen Maß mitwachsen kann, braucht es dringend innovative Assistenzsysteme für effiziente, faire und rechtssichere Verfahren.

Ziele und Innovationsgehalt

Ziel von VERITAS ist die Entwicklung des KI-Assistenzsystems „ALIGN“. Dieser interaktive Copilot entlastet AnwenderInnen, indem er Akten vorstrukturiert und automatisiert auf ort- und zeitbezogene Widersprüche hinweist. Die Innovation liegt im hybriden, neuro-symbolischen Ansatz: Statt auf fehleranfällige "Black-Box"-Sprachmodelle (LLMs) zu setzen, nutzt VERITAS eine offene Multi-Agenten-Architektur. KI-Agenten extrahieren Fakten und überführen sie in einen deterministischen "Knowledge Graph". Darauf aufbauend decken analytische Agenten mittels logischer Regelwerke harte Widersprüche (z.B. physikalisch unmögliche Reiserouten) überprüfbar auf. VERITAS verankert "Compliance & Ethics by Design" (AI Act, DSGVO) direkt in der Architektur. Das Human-in-the-Loop-Prinzip wird via Explainable AI (XAI) im Interface erzwungen, um blindes Vertrauen in die Maschine ("Automation Bias") zu verhindern. Zur Wahrung der Datensouveränität läuft das System rein On-Premise auf lokaler Hardware.

Angestrebte Ergebnisse und Erkenntnisse

Zentrales Ergebnis ist ein funktionsfähiger, evaluierter Software-Prototyp für die Pilot Studie am BFA. Das System

demonstriert seine Leistung anhand zweier Orchestrator-gesteuerter Workflows:

1. Identifikationsklärung: KI-Abgleich von Stammdaten und Biometrie zur Aufdeckung von Identitätstäuschungen.
2. Qualitätssicherung des Bescheids: Fallübergreifende Konsistenzprüfung, die Einvernahmen mit Beweisen, Länderberichten, Weisungen und Judikatur (RIS) logisch abgleicht.

Die Evaluierung soll belegen, dass das System die Verfahrensdauer verkürzt, kognitiv entlastet und die Bescheidqualität sowie Fairness für AsylwerberInnen steigert. Langfristig liefert VERITAS ein modulares Open-Source-Framework als Blaupause für den rechtsstaatlichen KI-Einsatz in der europäischen Verwaltung (GovTech).

Abstract

VERITAS (Trustworthy Identity Verification through Transparent Agent Systems / Vertrauenswürdige Identitätsfeststellung durch Transparente Agenten-Systeme)

Initial Situation, Problem, and Motivation

The asylum system is characterized by complex, time-critical, and data-intensive procedures. Caseworkers at the Federal Office for Immigration and Asylum (BFA) evaluate a daily flood of unstructured information (identity documents, multi-hour interview transcripts, country reports, historical files). Deceptions regarding identity, age, and origin represent enormous challenges. Until now, cross-source consistency checks have mostly been carried out manually. Given the growing data volumes and strict legal deadlines (max. 6 months), this is cognitively extremely demanding and prone to errors. Since staffing levels cannot grow at the same rate, there is an urgent need for innovative assistance systems to ensure efficient, fair, and legally secure procedures.

Goals and Degree of Innovation

The goal of VERITAS is the development of the AI assistance system "ALIGN". This interactive copilot relieves users by pre-structuring files and automatically pointing out spatial and temporal contradictions. The innovation lies in the hybrid, neuro-symbolic approach: Instead of relying on error-prone "black-box" language models (LLMs), VERITAS uses an open multi-agent architecture. AI agents extract facts and transfer them into a deterministic "Knowledge Graph". Building on this, analytical agents use logical rule sets to verifiably uncover hard contradictions (e.g., physically impossible travel routes). VERITAS anchors "Compliance & Ethics by Design" (AI Act, GDPR) directly into its architecture. The human-in-the-loop principle is strictly enforced via Explainable AI (XAI) within the interface to prevent blind trust in the machine ("automation bias"). To preserve data sovereignty, the system runs entirely on-premise on local hardware.

Intended Results and Findings

The central result is a functional, evaluated software prototype for the pilot study at the BFA. The system demonstrates its performance through two orchestrator-controlled workflows:

1. Identity Clarification: AI-driven cross-checking of master data and biometrics to detect identity deception.
2. Quality Assurance of the Decision: Cross-case consistency checks that logically compare interviews with evidence, country

reports, internal directives, and jurisprudence (RIS).

The evaluation aims to prove that the system shortens processing times, provides cognitive relief, and increases decision quality as well as fairness for asylum seekers. In the long term, VERITAS delivers a modular open-source framework as a blueprint for the rule-of-law-compliant use of AI in European public administration (GovTech).

Projektkoordinator

- AIT Austrian Institute of Technology GmbH

Projektpartner

- Bundesministerium für Inneres
- Research Institute AG & Co KG
- X-Net Services GmbH