

Defence-related Research Action - DEFRA

ACRONYM: SIMBA

Titel: Smart Interactive Multi-Agent Bias-Resilient Analysis

Duur van het project: 1/12/2025 - 1/03/2029

Budget: 1.619.000 €

Trefwoorden: Inlichtingenanalyse, Buitenlandse informatiemaniplatie en inmenging (FIMI), Mensgerichte verklaarbare AI, Multi-agent AI (MAAI), Grote taalmodellen (LLM's), Desinformatie.

waarvan RHID bijdrage: 1.210.000 €

PROJECTBESCHRIJVING

Context: De inlichtingenomgeving stelt de Belgische Defensie voor aanzienlijke uitdagingen, gekenmerkt door een overweldigend volume aan informatie dat vaak wordt gecompromitteerd door geavanceerde desinformatie en hybride oorlogvoeringstactieken. Inlichtingenanalisten moeten in dit complexe landschap geloofwaardige dreigingen onderscheiden om kritieke besluitvorming te onderbouwen. Bestaande analytische tools richten zich vaak op analyse aan de oppervlakte en slagen er niet in de onderliggende argumentatiestructuren en potentiële biases (vooringenomenheden) binnen inlichtingenbronnen aan te pakken. Dit maakt Defensie kwetsbaar voor gebrekkige inlichtingen die gebaseerd zijn op manipulatie of cognitieve biases. Het SIMBA-project pakt deze kritieke leemte rechtstreeks aan, een behoefte die formeel is erkend en wordt ondersteund door de Algemene Dienst Inlichting en Veiligheid (ADIV/SGRS).

Algemene doelstellingen: Het hoofddoel van het project is om de capaciteiten van inlichtingenanalisten aanzienlijk te verbeteren door een innovatief systeem te ontwikkelen dat helpt bij het identificeren en evalueren van biases en logische drogredenen in tekstuele inlichtingenbronnen. Om dit te bereiken zal SIMBA:

- Een flexibel multi-agent AI (**MAAI**) raamwerk ontwikkelen dat in staat is meerdere gespecialiseerde software-agenten te integreren die zich richten op bronanalyse en de detectie van cognitieve biases en logische drogredenen.
- Geavanceerde mogelijkheden voor natuurlijke taalverwerking (**NLP**) integreren door het gespecialiseerde Europese taalmodel van Textgain, CaLICO, aan te passen om de analytische agenten aan te sturen.

- Een adaptieve, mensgerichte interface ontwerpen op basis van 'Explainable AI' (**XAI**) principes om bevindingen op een duidelijke, intuïtieve en actiegerichte manier aan analisten te presenteren.
- Alle componenten integreren in een functionele systeemdemonstrator (**TRL 5-6**) en de capaciteiten valideren in realistische, gesimuleerde scenario's voor inlichtingenanalyse die relevant zijn voor Defensie.

Methodologie: Het 36 maanden durende project zal worden uitgevoerd via een samenwerkingsverband tussen de KU Leuven (coördinatie en XAI-interface), Sopra Steria (MAAI-raamwerk en systeemintegratie) en Textgain (NLP/LLM-capaciteiten). De methodologie hanteert een iteratieve, gebruikersgerichte aanpak, waarbij via regelmatige feedbackloops wordt gezorgd voor een continue afstemming op de behoeften van inlichtingenanalisten. De ontwikkeling verloopt in drie hoofdfasen: fundering en ontwerp (M1-M9), kernontwikkeling en integratie (M10-M24), en evaluatie, verfijning en einddemonstratie (M25-M36).

Potentiële impact van het onderzoek op de Belgische Defensie: SIMBA heeft als doel de snelheid, nauwkeurigheid en betrouwbaarheid van de inlichtingencyclus aanzienlijk te verhogen. Door de detectie van potentiële redeneerfouten en biases semi-automatisch te laten verlopen, stelt het systeemanalisten in staat zich te concentreren op beoordeling op hoger niveau en kritisch denken. Dit heeft een directe impact op Defensie door het verbeteren van het situatiebewustzijn (*situational awareness*), het ondersteunen van robuustere besluitvorming en het versterken van de weerbaarheid tegen informatiemaniplatie en dreigingen van cognitieve oorlogvoering. Het project is ontworpen voor mogelijke toekomstige integratie in de 'Next-generation AI-supported Integrated Intelligence Software Suite' (NAIIS) van de ADIV, wat de praktische relevantie waarborgt en de Belgische expertise in geavanceerde AI-toepassingen voor defensie en veiligheid versterkt.

Beschrijving van de verwachte onderzoeksresultaten en valorisatieperspectieven: Het belangrijkste verwachte resultaat is een **TRL 5-6** demonstrator van het SIMBA-systeem, gevalideerd in twee defensiespecifieke gebruiksscenario's (buitenlandse beïnvloeding met betrekking tot het conflict tussen Oekraïne en Rusland, en NAVO-perceptie). Een derde, publiekgerichte demonstrator zal een 'dual-use' toepassing tonen ter ondersteuning van fact-checking.

Andere belangrijke resultaten zijn onder meer:

- Een herbruikbaar MAAI-raamwerk voor bias-bestendige analyse.
- Nieuwe NLP/LLM-technieken voor de detectie van biases en drogredenen.
- Gevalideerde mensgerichte XAI-methoden en componenten voor gebruikersinterfaces.
- Diverse wetenschappelijke publicaties (peer-reviewed).

Op korte termijn is de Belgische Defensie de primaire gebruiker, met potentieel voor integratie in NAIIS en gebruik in de opleiding van analisten. Op middellange termijn zullen de projectpartners de resultaten valoriseren via verbeterde commerciële diensten (Sopra Steria, Textgain) en verder academisch onderzoek (KU Leuven). Het project legt ook de basis voor vervolginiciatieven die erop gericht zijn de technologie operationeel te maken (TRL 7-9) of toe te passen in nieuwe domeinen zoals wetshandhaving en crisisbeheer.

CONTACT INFORMATIE

Coordinator

Katrien Verbert
KU Leuven – Department of Computer Science
katrien.verbert@kuleuven.be

Partners

Cédric Genin
Sopra Steria Belgium – Defence & Security Business Unit
cedric.genin@soprasteria.com

Redouan el Hamouchi
Textgain
redouan@textgain.com

LINK(S) NAAR PROJECT

(Nog) Niet beschikbaar.