

KUDZAI SAUKA

Ph.D. Candidate | AI Transparency, Explainability & Conversational AI

+31 641 065 646 k.sauka@hva.nl ksauka.github.io github.com/ksauka [LinkedIn](#) [ORCID](#)

RESEARCH SUMMARY

PhD Researcher and Research Scientist at the Hogeschool van Amsterdam, pursuing a doctorate at the University of Amsterdam (Faculty of Business and Economics). My research develops **layered explainability architectures for conversational AI**, combining Dempster-Shafer Theory (DST) for belief-guided intent clarification with Knowledge Graph-Augmented Retrieval (KG-RAG) and counterfactual faithfulness auditing. I apply a unified evaluation spine across all work — assessing *faithfulness* (perturbation-based), *usability* (verification accuracy and time), and *reliance calibration* (appropriate trust under transparency and anthropomorphic framing). My work spans XAI, conversational agents, semantic web technologies, ontology engineering, and human-centred AI evaluation. Looking ahead, I am extending this work into **agentic AI settings**, framing DST belief-modeling as infrastructure for agent uncertainty tracking in multi-agent pipelines.

EDUCATION

Ph.D. Candidate: AI Transparency and Explainability **June 2023 — Present**
Faculty of Business and Economics, University of Amsterdam, Netherlands.

THESIS (WORKING TITLE): *Human-Centred Explainable Conversational AI: Belief-Guided Clarification and Evidence-Faithful KG-RAG*

RESEARCH AREAS: Explainable AI (XAI) · Conversational Agent Transparency · Dempster-Shafer Theory for Intent Disambiguation · Knowledge Graph-Augmented Retrieval (KG-RAG) · Counterfactual Faithfulness Auditing (SMILE-CF) · Trust & Reliance Calibration · Anthropomorphism in AI · Semantic Web & Ontology Engineering · Human-Centred Evaluation

Master of AI Software **Mar 2021 — Feb 2023**

THESIS: *Adversarial Robust Evaluation Metric Formulation Through Explainable AI and Adversarial Training: A Case for Network Intrusion Detection Systems.*

GPA: 4.39/4.50 · Dept. of AI Software, Gachon University, Republic of Korea.

Master of Global Business and Entrepreneurship — Finance Major **Mar 2019 — Feb 2021**

THESIS: *The Dynamics of Sub-Saharan Africa Non-Performing Loans and Macroeconomic Effects: Evidence from Oil Exporting and Non-Oil Exporting Countries.*

GPA: 4.50/4.50 · Dept. of Global Business and Entrepreneurship, Gachon University, Republic of Korea.

B.Tech (Hons) Financial Engineering **Aug 2012 — June 2016**

THESIS: *Modelling the Sensitivity of Zimbabwean Commercial Banks' Non-Performing Loans to Shocks in Macro- and Micro-Economic Variables (2009–2014).*

GPA: 3.3/4.00 · Dept. of Business Management Science, Harare Institute of Technology (HIT), Zimbabwe.

PROFESSIONAL EXPERIENCE

Research Scientist — Explainable Conversational AI **June 2023 — Present**
Centre for Market Insights (CMI) Lab *Hogeschool van Amsterdam, Netherlands*

- Developing layered XAI architectures for conversational agents: DST-based belief-guided intent clarification (Layer 1) and KG-conditioned keyphrase rationales with SMILE-CF faithfulness auditing (Layer 2).
- Designing and running human-subject studies on transparency, anthropomorphism, and trust calibration in AI systems (vignette experiments, fsQCA, SPSS).
- Building *AnthroKit*: a reusable design system for controlled anthropomorphic tone manipulation in experimental chatbot research.
- Supervising and tutoring in Online Data Mining.
- Active publication record across XAI, conversational AI, knowledge graphs, and AI-entrepreneurship (10 papers; 4 accepted, see Publications).

AI Security Researcher **Mar 2021 — Mar 2023**
Information Security Lab (ISLAB) *Gachon University, Republic of Korea*

- Developed explainable deep learning-based Network Intrusion Detection Systems (NIDS).
- Designed adversarially robust NIDS using Bayesian Stackelberg game-theoretic formulations.
- Published adversarial robustness and explainability research in Q2 journals and international conferences (ISIS 2021, KIIS 2022).

Business Statistics Teaching Assistant

Mar 2019 — Feb 2020

Corporate Finance Research Assistant

Mar 2020 — Feb 2021

Dept. of Global Business and Entrepreneurship

Gachon University, Republic of Korea

- Developed deep-learning financial forecasting models for stock price prediction.
- Applied advanced dynamic panel data models (GMM, fixed/random effects) for macroeconomic analysis.
- Conducted undergraduate Business Statistics lectures, grading, and exam invigilation.

Finance and Administration Officer

Oct 2018 — Feb 2019

ReiMac Trading (PVT) Ltd

Harare, Zimbabwe

- Budget preparation, financial record management, and daily/monthly transaction reconciliation.

Tutor — Business Calculations & Quantitative Techniques

July 2018 — Jan 2019

SPECIS College, School of Business

Harare, Zimbabwe

- Delivered LCCI Advanced Business Calculations and IBAS Quantitative Techniques lectures; administered grading and projects.

Finance and Administration Officer

Aug 2014 — Sept 2018

Housi6 (formerly Morden Brick Properties)

Harare, Zimbabwe

- Maintained accounts payable/receivable systems; served as cross-departmental financial liaison; resolved billing disputes.

PUBLICATIONS

JOURNAL ARTICLES

1. Wang, Y., **Sauka, K.**, & Situmeang, F. B. I. (2025). Anthropomorphism and transparency interplay on consumer behaviour in generative AI-driven marketing communication. *Journal of Consumer Marketing*, 42(4), 512–536. [H-Index: 122 · Q1: Business, Management & Accounting]
2. **Sauka, K.**, Shin, G.-Y., Kim, D.-W., & Han, M.-M. (2022). Adversarial robust and explainable network intrusion detection systems based on deep learning. *Applied Sciences*, 12(13), 6451. doi:10.3390/app12136451 [Impact Factor: 2.838 · Q2: Computer Science Applications]
3. Choi, S. B., **Sauka, K.**, & Lee, M. (2024). Dynamic capital structure adjustment: An integrated analysis of firm-specific and macroeconomic factors in Korean firms. *International Journal of Financial Studies*, 12, 26. doi:10.3390/ijfs12010026 [Impact Factor: 2.3 · Q2: Finance]
4. Choi, S. B., **Sauka, K.**, & Lee, M. (2024). Dynamics of dividend payout in Korean corporations: A comprehensive panel analysis across economic cycles. *Journal of Risk and Financial Management*, 17(12), 564. doi:10.3390/jrfm17120564 [CiteScore: 4.5 · Q2: Finance · Q1: Business, Management & Accounting]
5. Muvingi, J., **Sauka, K.**, Chisunga, D., & Chirume, C. (2017). Modelling the sensitivity of Zimbabwean commercial banks' non-performing loans to shocks in macro- and micro-economic variables (2009–2014). *International Journal of Economic Behavior and Organization*, 5(4), 92–99. doi:10.11648/j.ijebo.20170504.12

CONFERENCE PAPERS

1. **Sauka, K.**, Saou, Y., Situmeang, F. B., & Kackovic, M. (2025). A nexus of explainability and anthropomorphism in AI-chatbots. In *World Conference on Explainable Artificial Intelligence* (pp. 115–138). Cham: Springer Nature Switzerland.
2. **Sauka, K.** (2025). Interactive intent classification with belief-based proactive clarification. In *Proceedings of the ProActLLM Workshop, CIKM 2025*. [Accepted]
3. **Sauka, K.** (2025). Building hierarchy-aware knowledge graphs (KPN). In *ISWC 2025 Industry Track*. [Accepted]
4. **Sauka, K.**, Hong, S.-S., & Han, M.-M. (2021). Adversarial robust and explainable network intrusion detection systems based on deep learning. *ISIS 2021 — 22nd International Symposium on Advanced Intelligent Systems*, Dec. 15–18, 2021, Cheongju, South Korea.
5. **Sauka, K.**, Shin, G.-Y., Kim, D.-W., & Han, M.-M. (2022). Bayesian Stackelberg adversarial games towards robust deep learning-based network intrusion detection systems. *KIIS Autumn Conference 2022*, Vol. 32, No. 2, Oct. 28–29, 2022.

WORKING PAPERS & MANUSCRIPTS IN PREPARATION

1. **Sauka, K.** et al. (2026, target). AnthroKit: A reusable design system for anthropomorphic chatbot research.
2. **Sauka, K.** (2026, target). Belief-guided proactive clarification in dialogue: A DST extension for intent uncertainty tracking.
3. **Sauka, K.** (2026, target). Designing for trust: KG-conditioned explanations for accountable conversational AI ecosystems.
4. **Sauka, K.** (2026, target). Beyond correct citations in KG-RAG: SMILE-CF counterfactual faithfulness auditing.
5. **Sauka, K.** et al. (2026, target). AI & entrepreneurship: A bibliometric analysis of a decade of research.

PROJECTS

- **DSAgent** — DST-based intent classification system across ATIS, SNIPS, CLINC150, BANKING77, TOPv2 datasets. Full hierarchical belief logging (cosine similarity, mass functions, belief updates, combined mass) per query; JSON log output. Python / scikit-learn.
- **AnthroKit** — Reusable design system for anthropomorphic tone manipulation in conversational AI research. Big Five personality mapping with bounded presets for experimental comparability. Deployed in a loan pre-assessment XAI chatbot.
- **ESD Platform** — Flask backend with Neo4j knowledge graph, GraphRAG pipeline, and modular architecture for explainable semantic search.
- **Nonperforming Loans Prediction System** (Capstone) — Macro/micro-economic NPL prediction model. Harare Institute of Technology, Zimbabwe, 2015–2016.

SKILLS

Programming	Python, R
Statistical Tools	STATA, EViews, SPSS, VOSviewer
ML / DL Libraries	PyTorch, scikit-learn, NumPy, SciPy, Pandas, Matplotlib/Seaborn
XAI & Reasoning	SHAP, LIME, Dempster-Shafer Theory (DST), Counterfactual Explanation, SMILE-CF
NLP & Conv. AI	LangChain, LangGraph, intent classification, dialogue state tracking, LLM prompting
Knowledge Graphs	Neo4j, RDF/OWL ontology engineering, GraphRAG, OPLa modular ontology design
Security / Adv. ML	Adversarial Toolbox (ART), CleverHans, Bayesian game-theoretic adversarial modelling
Web & Backend	Flask, FastAPI, GitHub Pages, HTML/CSS/JS
Research Methods	Vignette experiments, fsQCA, dynamic panel data (GMM), bibliometric analysis
Authoring	LaTeX / Overleaf, Git / GitHub