Ayman Saeed

Ayman Saeed | aymans@aims.ac.za | +249 906099000 +27 797876235

I am passionate about machine learning, with a strong background in mathematics and programming, and hands-on experience in building and understanding deep learning models. My areas of interest include Natural Language Processing (NLP), model interpretability, and AI Safety.

Education:

Sep/2024 – Jul/2024	**Master's in Artificial Intelligence for Science (AI for Science)**
Cape Town, South Africa	University of Cape Town / African Institute of Mathematical Sciences
Oct/2016 – Nov/2023	**Bachelor of Science in Electrical and Electronic Engineering**
Khartoum, Sudan	University of Khartoum
Grade	9.03 out of 10 (Ranked **1st** in class).
Comment	I completed my degree in 7 years because of the political strikes in Sudan.

Experience:

AI & NLP Expert (AI Research Foundations – Train-the-Trainer) | AIMS South Africa (Nov/2025 – Present):

- Deliver Train-the-Trainer sessions for HEI lecturers ("AI Champions") and teaching assistants, enabling local delivery of advanced AI education across Nigeria, Ghana, Kenya, and South Africa.
- Support curriculum localisation, quality assurance, and ongoing technical mentoring for capstone/research projects.
- Programme delivered by AIMS South Africa in partnership with FATE Foundation, with support from Google.org.
- Curriculum is built around Google DeepMind's AI Research Foundations (developed with UCL).

Teaching Assistant | AIMS South Africa (Aug/2025 – Present):

- Provide academic support to master's students, assist with tutorials, and grade assignments for graduate-level courses.

Research & Projects:

Benchmarking Interdependent Privacy in AI Assistants (Master's Research, Jan/2025 –June/2025):

- Objective: Evaluate how well-advanced AI assistants handle interdependent privacy scenarios, where the private information of others is affected by a user's disclosure.
- Benchmarking Framework: Extend the CI-Bench evaluation to capture multi-party privacy risks, with new synthetic test cases reflecting family, health, and social contexts.
- Methodology: Design interdependent privacy scenarios using structured representations of context, actors, and information flows. Generate realistic multi-turn scenarios, then evaluate AI assistant behavior using state-of-the-art language models (e.g., GPT, Gemini) on tasks such as context interpretation, norm identification, and appropriateness judgment.
  'Master's research supervised by Prof. Ulrich Aïvodji (MILA/ÉTS).'

Model Extraction Attacks on DistilBERT (2022-2023):

- Objective: Investigate model extraction attacks on DistilBERT to steal its functionality by training a substitute model.
- Key Findings: Pre-trained models outperformed those trained from scratch, with TinyBERT surpassing BERT-small in performance despite having fewer parameters, due to its training with Knowledge Distillation. Results also emphasized the impact of data similarity and volume on extraction effectiveness.
- Impact: Demonstrated security vulnerabilities in DistilBERT, emphasizing the need for robust defenses against model extraction attacks.

Publications:

- Submitted Paper with the title "Model Extraction Attacks on DistilBERT" on The International Conference on Learning Representations (ICLR2023), and it got accepted to be one of the archived papers in their database.

Achievements:

- Academic Excellence Award (AIMS South Africa): Awarded to the top three highest academic achievers in the "AI for Science" master's program at AIMS South Africa.
- Best Graduation Project (2023): This is a yearly prize awarded by the University of Khartoum to the creators of the best graduation project in the Engineering Faculty. This award was given to me and my partner in the year 2023.
- University Excellence Award: I was honored by the Dean of the Engineering College at the University of Khartoum for my outstanding academic achievements, attaining the highest position in my class with a CGPA of **9.03**/10

Research Interest:

Natural Language Processing, Large Language Models, AI Safety, models interpretability

Skills:

Programming & Development

- Languages: Python, C++, C, MATLAB, Java.
- Frameworks: PyTorch, TensorFlow, JAX, NumPyro.
- Software Engineering: Git, Docker, Bash.

Computer Vision & NLP

- Vision: Classification, Object Detection, OpenCV.
- NLP: Tokenization, Pretrained Transformers

Competitive Programming

- Algorithmic Thinking, Data Structures, Problem-Solving.

Machine Learning & Deep Learning

- Neural Networks: CNNs, BNNs, Transformers.
- Uncertainty Estimation: MC Dropout, Deep Ensembles, Variational Inference.
- Foundation Models: LLMs, TabPFN.
- Generative AI: VAEs, GANs, Diffusion Models, Probabilistic Models
- RL: PPO, Value Iteration, Policy Gradients, RLHF
- Fairness, Robustness & Interpretability
- Active Learning & Data Efficiency
- Math. Foundations: Information Theory, Variational Methods

Conferences:

Deep Learning Indaba AI Conference, Tunisia (2022), Rwanda (2025):

Deep Learning Indaba is an annual event aimed at promoting machine learning and AI across Africa, lecturing workshops, talks, and networking opportunities for the AI community to foster research and innovation.

Middle East & North Africa Machine Learning (MENAML) Winter School, Qatar (2025), Saudi (2026):

A program focused on advancing machine learning research and applications in the MENA region. Participated in lectures and hands-on workshops led by leading AI researchers. Competed in a (CLV) Prediction Challenge, where I secured 2nd place.

Indaba-X AI conference, Khartoum (2021,2022):

Co-organized the biggest AI conference in Sudan twice. Moreover, worked as a mentor explaining mathematical and machine learning concepts in the 2021 version and volunteered as a tutor with a lecturer in explaining basic Natural Language Processing (NLP) concepts in the 2022 version.

Volunteers:

- Member at Black In AI.
- Member of the Electrical and Electronic Engineering Students Exhibition (EEESE) academic committee (Nov/2018 – Mar/2020).