Asim Osman

Cape Town, South Africa | asim.awad98@gmail.com | Github | Portfolio | +27 76 425 6768

Profile

Research engineer and Google DeepMind Scholar with an MSc in AI (Distinction, top of cohort) from the University of Cape Town / AIMS South Africa. I'm focused on reinforcement learning and LLM systems. My most recent works include new contrastive multi-agent RL algorithm at InstaDeep, and an LLM agent for ML engineering in my MSc thesis. My interests span reinforcement learning (MARL), LLM agents and reasoning, and the systems side of ML; LLM serving and distributed training (sharding, multi-device). I'm looking for ML / research-engineering roles in reinforcement learning, LLM Agentic Systems, reasoning and inference strategies, or ML infrastructure.

Experience

Research Engineer Intern *Aug 2025 – Present*
InstaDeep, Cape Town, South Africa

- Worked on developing **Tinkerer**, an autonomous LLM Agent for automated AI research (idea generation → code → experiment scheduling → result collection).
- Developed **CPPO** (and earlier **Contrastive PQN**), the first on-policy contrastive RL algorithm — reward-free goal-conditioned learning that matches or beats reward-engineered PPO across 18 benchmark tasks. Preprint: arXiv:2605.13554; under review at NeurIPS 2026.

Research Engineer Intern *May – Aug 2024*
Enigma AI R&D, Khartoum, Sudan

- Worked on Field-Monitoring App, an interactive web app that pulls satellite imagery and maps field boundaries (Mapbox/GeoJSON) to track crop and field health over time.

Data Science Intern *Jul 2022 – Mar 2023*
AmunData, Khartoum, Sudan

- Worked on to agriAI farming assistant and crop-classification pipeline (Random Forest + Sentinel-2 multispectral bands), reaching 89.22% accuracy across key crops in El Gezira.

Education

MSc in Artificial Intelligence (AI for Science) *2024 – 2025*
University of Cape Town (UCT) / African Institute for Mathematical Sciences (AIMS) South Africa

- Graduated with **Distinction** — AIMS Excellence Award (top student in cohort); Google DeepMind Scholarship recipient. *Transcript*
- Thesis: *Scaling Inference-Time Compute for Machine Learning Engineering Agents*
  Supervisor: **Arnu Pretorius** (Stellenbosch University)
- Distinction across 15+ courses incl. Computer Vision, Bayesian Inference, CUDA, RL, LLMs, Deep Generative Models, and Applied ML at Scale.

BSc (Hons) in Electrical & Electronic / Software Engineering *2016 – 2023*
University of Khartoum, Sudan — Certificate & Transcript

- Final GPA **9.18 / 10.0**; Best Graduation Project Award.
- Thesis: Benchmarking Neural Machine Translation for Arabic–Swahili; released the largest Arabic–Swahili translation dataset. Published at ICLR 2023, Tiny Papers.

Publications

- Asim Osman, et al. *Self-Supervised On-Policy Reinforcement Learning via Contrastive Proximal Policy Optimisation (CPPO).* arXiv:2605.13554, 2026. Under review at NeurIPS 2026; submitted to the RLBRew Workshop, RLC 2026.
- Asim Osman, Ahmed Elmahdy, Hiba Imam. *Machine Translation Baselines for Arabic–Swahili.* ICLR 2023, Tiny Papers Initiative.

Selected Projects

CPPO: Contrastive Proximal Policy Optimisation | Paper | Code
The first on-policy contrastive RL algorithm, learns goal-conditioned Q-values via contrastive learning and derives the policy advantage directly from those contrastive Q-values removing the separate value network and GAE entirely.
Key results and contributions:

- Implemented using JAX/Flax framework with feedforward + recurrent variants, and works for single and multi agent RL (**CPPO and ICPPO**) I also developed Contrastive PQN a value based variant and added automatic curriculum learning mechanism on top of the contrastive RL systems developed.
- Evaluated across **18 tasks** in 5 suites (Navix, JaxGCRL, SMAX, Connector, JaxNav), this algorithm matches or exceeds dense reward PPO on **12/18** tasks and outperforms off-policy contrastive-RL baselines on **14/18**.

AIDE Kaggle Agent: AI-Driven Exploration for ML Engineering |Code | Thesis
An open-source, autonomous agent that solves end-to-end ML Kaggle competitions by framing development as a *solution space tree search*. Powered by Open source LLMs, AIDE drafts initial solutions, debugs code, and iteratively improves working scripts—mirroring the workflow of a human data scientist.
Key elements and results:

- Implements inference-time scaling (ITS) strategies: *self-consistency*, *self-reflection*, and *decomposed planner–coder (task decomposition)* to enhance smaller, open-source models.
- Self-hosted a high-throughput `vLLM` backend (replacing AIDE's API backend) serving distilled DeepSeek-R1 models (7B/14B/32B), running up to 10 parallel agent processes on one H100 with per-process memory budgeting and CPU pinning; evaluated on a curated subset of MLE-Bench competitions.
- The best agent—**32B + Decomposed Planner–Coder**—achieved a **30%** medal rate, *matching* OpenAI `o4-mini` and *surpassing* the GPT-4-Turbo baseline on our subset.

Fine-Grained Credit Assignment for RLHF (VinePPO) — Code
A JAX-based RL framework that assigns token-level rewards to stabilize LLM alignment training.
Key contributions:

- Migrated the training pipeline from PyTorch to JAX/Flax, with TPU-optimized training.
- Built modular, API-based credit assignment using GPT-4 / Gemini as judges.

Arabic–Swahili Neural Machine Translation — Paper (ICLR 2023) | Code
First large-scale Arabic–Swahili NMT study for a low-resource language pair.
Key contributions:

- Created the first large-scale Arabic–Swahili parallel corpus (around 30k verse-aligned sentence pairs).
- Built bidirectional baselines: a Transformer trained from scratch (Fairseq) plus fine-tuned **M2M-100** and **mT5** (Fairseq + HuggingFace).
- Showed that fine-tuning a large multilingual model on a small dataset significantly outperforms training from scratch for this morphologically rich pair.

Skills

Programming: Python, JAX/Flax, PyTorch, NumPy, Pandas, C/C++ (CUDA), Bash, Git

Reinforcement Learning: contrastive RL (CRL), goal-conditioned RL, PPO, multi-agent RL (IPPO/MAPPO), curriculum learning (SFL), RLHF (PPO/GRPO/DPO)

LLMs & NLP: Transformenrs and Attention, Language Modelling, LLM agents, inference-time scaling, LLM fine-tuning (SFT/LoRA/RLHF), vLLM serving, HuggingFace, TRL, LangGraph

ML Systems & MLOps: GPU/TPU training, parallel inference, distributed-training concepts (sharding/FSDP), Docker, Hydra, experiment tracking (W&B, MLflow, Neptune), cloud (AICHOR/AWS/GCP)

Deep Learning & Machine Learning: computer vision (CNNs, transfer learning), generative AI, Bayesian deep learning, active learning, information theory, classical ML (scikit-learn)

Reference
Sasha Abramowitz — Research Engineer, InstaDeep
s.abramowitz@instadeep.com