White Paper: Crane Models – Pioneering Sovereign-Grounded AI for Uganda

A Vision for Sovereign, Relevant, and Responsible Artificial Intelligence

Industry 4.0 Bureau | STI - Office Of The President
AI Studio Uganda, DeepTech Center Of Excellence

Kampala, Uganda | April 2025

Bronson Bakunga, Kato Steven Mubiru, Kateeba David, Kirabo Melissa, Kavuma Lameck, Bujjingo Treasure, Roland Ganafa, Sibomana Glorry, Lwanga Caleb, David Gonahasa, Gimei Alex, Wesley Kambale , Mutegeki Henry, Atuhaire Collins, JohnRoy Nangeso, Tukamushaba Catherine et al.

Version 1.3 - Enhanced Public Release

Executive Summary

The rapid advancement of Artificial Intelligence (AI), especially Large Language Models (LLMs), offers unprecedented opportunities globally (Stanford HAI AI Index, 2024). For Uganda, this presents a unique moment to leverage AI for social economic transformation, but realizing this potential requires AI systems incorporating the nation's rich cultural and linguistic diversity. Standard global LLMs often lack this local grounding due to their training origins (Bender et al., 2021), necessitating tailored solutions. Project Crane, initiated by the STI Secretariat – Office of the President, Uganda (Industry 4.0 Bureau), takes advantage of this opportunity, focusing on local ownership, data sovereignty, and culturally tuned model development.

Driven by the AI Studio Uganda (DeepTech Center of Excellence) and comprised of young Ugandan engineers working together with local experts and global partners. Project Crane is developing Crane Models, Uganda's first sovereign family of LLMs. Our mission is to engineer AI that authentically understands and interacts within Uganda's diverse landscape. Our approach integrates Cultural Grounding, Resource-Efficient Technical Innovation, and Sovereign Deployment. We adapt state-of-the-art foundation models (commencing with Google's Gemma family (Google LLC, 2025)) using a novel synthetic data pipeline that is inspired by global best practices (Liu et al., 2024), critically validated by a growing number of Ugandan elders and cultural experts to ensure alignment with local knowledge. This will be deployed securely on locally accessible infrastructure (i.e., Afriqloud), which reinforces Uganda's commitment to AI empowerment, ownership, data security, and provides a reproducible blueprint for other African nations.

Central to our work is the Ugandan Cultural Context Benchmark (UCCB) Suite, the first comprehensive framework for evaluating AI on Ugandan cultural knowledge (UCCB-Knowledge) and nuanced social reasoning/bias sensitivity (UCCB-Nuance). Developed with a growing number of Ugandan domain experts from institutions such as Makerere University, Kyambogo University, and aligned with global benchmark quality standards (Reuel et al., 2024), UCCB enables rigorous assessment.

The first model, Crane Gemma 4B, is currently undergoing pre-training and fine-tuning. Initial internal projections anticipate significant gains over baseline models, promising improved understanding vital for applications in education (e.g. accurate UNEB context understanding), agriculture, and cultural heritage preservation. Our integrated security framework incorporates input/output safeguards inspired by tools like GemmaShield or LlamaGuard, acknowledging potential interoperability challenges between different ecosystems . Security is enhanced through rigorous adversarial testing via Microsoft's PyRIT (Microsoft Corporation, 2024) and robust data access control via micro-segmentation on Afriqloud, leveraging Red Hat's Skupper project (Red Hat Skupper, 2025). This multi-layered approach targets >90% block rate on simulated attacks.

Project Crane leverages strategic technical collaborations with resources and professionals from Google DeepMind (Gemma models), Cohere (Aya & Command R+ models), IBM (Granite models, InstructLab methodology), Red Hat (OpenShift deployment expertise, Neural Magic inference acceleration), Afriqloud (sovereign infrastructure), Kove (memory virtualization technology), Weights & Biases (MLOps best practices), Hugging Face (dataset/model hosting standards), Alter Domus/Coinbase (AI security expertise), and the Enterprise Neurosystem (non-profit open source AI research community).This initiative positions Uganda at the forefront of developing equitable, culturally-attuned AI, contributing to Africa, open science, and supporting UN Sustainable Development Goals (UNDP). We invite global collaboration via our forthcoming open release of UCCB components at `https://huggingface.co/STIAIStudio`.

1 THE OPPORTUNITY: CULTURALLY-GROUNDED AI FOR UGANDA

1.1 The Global AI Landscape and the Cultural Divide

Artificial Intelligence (AI), particularly through powerful Large Language Models (LLMs), is rapidly reshaping industries and global information access (McKinsey & Company, 2023). These models exhibit advanced natural language capabilities, powering applications from sophisticated search to automated content creation. However, their pre-training corpora are predominantly sourced from North American and European data, embedding Western cultural norms, values, and linguistic patterns into model outputs (Bender et al., 2021). Consequently, this bias creates a significant cultural divide, limiting model effectiveness and relevance in diverse global contexts (VentureBeat, 2024). Benchmark evaluations such as BLEnD reveal substantial performance gaps, with models failing to grasp nuanced cultural references and everyday knowledge outside their primary training domains (Plewis et al., 2024). For instance, while capable of describing generic ceremonies, standard LLMs often misinterpret the sequence and significance of specific cultural rites such as Ankole's 'Kuhinjira, Buganda's Kwanjula, Busoga's kwadhura , Langi's Nyom, Bagisu's and Basamia's Ohwanjula, which mean introduction ceremony.As noted by Emily M. Bender, this lack of true understanding, despite predictive capabilities, poses risks of misinterpretation and bias propagation (Time, 2023). Furthermore, neglecting linguistic diversity impacts the fundamental robustness of AI itself. LLMs represent high-dimensional knowledge spaces (Aghajanyan, A & Zettlemoyer, 2020), and the failure to incorporate diverse linguistic data leaves vast dimensions unsaturated, increasing susceptibility to hallucinations and biased outputs (McKenna, N. , Li, T. & Cheng, 2023). Preserving indigenous languages and their unique reasoning patterns is therefore critical not only for cultural identity but also for the resilience and completeness of the global AI endeavor. The loss of a language represents the irreversible disappearance of a unique cognitive framework, akin to losing potential scientific breakthroughs hidden in biodiversity (Abrams & Strogatz, 2003). Counteracting 'language death' is essential for continued human and technological evolution, as the breadth and balance of training data directly impact model generalizability and reasoning capabilities.

1.2 Uganda's Unique Context: Challenges & Opportunities

Uganda presents a unique confluence of cultural richness and technological aspiration, making it fertile ground for pioneering culturally-adapted AI:

- **Linguistic Diversity:** Home to over 40 indigenous languages spanning Bantu, Nilotic, and Central Sudanic families, alongside Ugandan English and Swahili (Nekoto et al., 2020), this presents challenges for standard NLP models but drives the need for inclusive multilingual solutions.
- **Rich Cultural Heritage:** Diverse ethnic groups maintain unique traditions, social norms (e.g., specific forms of address for elders), artistic expressions (e.g., Larakaraka dance), historical narratives, and indigenous knowledge systems largely absent from global datasets.
- **Demographic Dividend:** With over 73% of its population under 30 and 14.2 million internet users as of early 2025 (DataReportal, 2025), Uganda has a vast, digitally-engaged youth demographic demanding relevant technology.
- **Growing Tech Ecosystem:** Innovation hubs such as National ICT Hub, the Innovation Village (Innovation Village, 2025) foster local talent and tech-driven solutions, providing a foundation for initiatives like Project Crane.
- **Critical Biodiversity:** Hosting ecosystems like the Bwindi Impenetrable Forest (UNESCO, 2025) necessitates innovative AI tools for environmental monitoring and conservation.

This multifaceted context renders generic AI solutions inadequate but provides the **’fuel of AI’**—a rich source of unique linguistic data and cultural knowledge. It creates a distinct opportunity to develop methodologies for Efficient Multilingual Model Development, focusing on maximizing generalizability across Uganda’s diverse landscape while minimizing trainable parameters, alongside cultural knowledge integration and context-specific reasoning applicable across Africa.

Figure: Africa with a zoomed-in view of Uganda.

1.3 Real-World Impact of Culturally Misaligned AI

Deploying culturally non-tuned AI in Uganda carries tangible negative consequences across vital sectors:

- **Education:** AI tutors risk supplying inaccurate information about Ugandan history (e.g., misstating kingdom formations) or misinterpreting literary texts, undermining national curricula and cultural identity (General LLM Inaccuracy, 2023).
- **Communication:** Translation tools or chatbots frequently misinterpret Ugandan English idioms (e.g., "to be tight" meaning "to be friends") or fail to grasp required levels of politeness in specific social interactions (ACL Anthology, 2023).
- **Agriculture:** Generic advisory systems can provide advice unsuitable for Uganda’s specific microclimates, soil types, or indigenous crops. Conversely, tailored mobile advisories have shown significant positive impact (UNFCCC, 2023).
- **Health:** AI health assistants lacking proficiency in local languages can impede effective patient communication and understanding of traditional health concepts, hindering care quality (Nilepost, 2023).
- **Identity & Representation:** The consistent absence or misrepresentation of Ugandan cultural content in global AI tools can negatively impact cultural confidence, particularly among youth (Bender et al., 2021).

1.4 Crane Models: A Vision for Relevant AI

Project Crane and its Crane Models family offer Uganda's strategic, locally driven response to these challenges, embodying a vision for AI that empowers:

- **Preserves & Promotes Heritage:** Aims to accurately document and enable interactive exploration of Ugandan folklore, proverbs (like those documented by Nekoto et al., 2020), languages, and traditions, creating accessible digital cultural resources.
- **Enhances Education:** Seeks to deliver reliable, engaging, and culturally validated information aligned with Ugandan curricula (UNEB and NCDC context) and knowledge systems, supporting dialect preservation and effective learning (Education Tech, 2023).
- **Boosts Local Innovation:** Empowers Ugandan developers, researchers (at hubs and centers like DeepTech Center of Excellence , National Innovation Hub e.t.c), and entrepreneurs with foundational models that understand local needs, fueling applications in key sectors.
- **Supports Key Economic Sectors:** Provides context-aware tools for agriculture (building on pilot success (Internal Pilot Data, 2023)) and tourism (accurate cultural site information).
- **Aligns with National Goals & Sovereignty:** Explicitly supports Uganda's initiatives such as Parish Development Model, Third National Development Plan (NDPIII) Vision 2040 , objectives on Agriculture, Tourism , Manufacturing and Science(ATMS). By embedding AI sovereignty through local development , deployment (Afriqloud), and governance, it ensures data control and that benefits are Ugandan-led.

2 TECHNICAL INNOVATION FRAMEWORK: BUILDING CRANE MODELS

Project Crane employs a multifaceted technical framework designed to address the unique challenges of developing culturally-grounded AI in Uganda. Our approach integrates rigorous benchmark development, innovative data strategies, efficient model adaptation, and a robust security architecture, all underscored by strong collaborative partnerships.

2.1 The Ugandan Cultural Context Benchmark (UCCB) Suite

Effective evaluation is paramount. We are developing the Ugandan Cultural Context Benchmark (UCCB) Suite as the cornerstone for assessing LLM capabilities within Uganda. This suite provides a standardized, culturally relevant measure of performance, guiding both our development and community efforts.

2.1.1 Design Philosophy & Structure

Inspired by the rigor of international benchmarks like C-Eval (Huang et al., 2023) and CBBQ (Chen et al., 2023), but tailored specifically for Uganda, UCCB comprises two main components:

- **UCCB-Knowledge:** Assesses factual understanding and basic reasoning across key Ugandan domains (History & Civics, Language & Communication, Arts & Culture, Social & Daily Life, General Knowledge, Multimodal Recognition) using primarily Multiple-Choice Questions (MCQs) and Instruction Following tasks across Basic, Intermediate, and Advanced difficulty levels. A UCCB-Knowledge-HARD subset targets advanced reasoning. Example Task: Identifying key figures from the Busoga Kingdom.

- **UCCB-Nuance:** Evaluates understanding of subtle cultural dynamics, social etiquette, value-based reasoning, and bias sensitivity (using paired ambiguous/disambiguous scenarios inspired by CBBQ) across categories like Cultural Bias, Social Interaction, and Complex Linguistic Interpretation. Example Task: Assessing the appropriateness of responses in a simulated traditional greeting scenario.

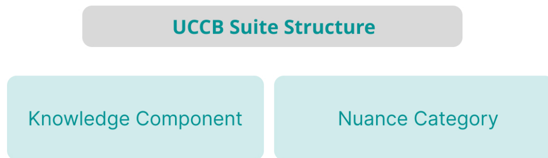

Figure 1: Conceptual layout of the UCCB Suite, dividing knowledge-based and nuance-based evaluation streams.

2.1.2 Development & Validation

UCCB development is fundamentally expert-led, involving Ugandan domain specialists from Makerere University, cultural institutions, and relevant sectors. Data is sourced from diverse Ugandan materials (archives, mock exams, local literature, validated public data) to prioritize authenticity and mitigate contamination (Huang et al., 2023). Initial benchmark items (targeting 5,000+) undergo multi-annotator validation and final expert review for cultural accuracy, clarity, and fairness. The entire process adheres to principles of high-quality benchmark design (Reuel et al., 2024).

2.1.3 UCCB Bench Categories

Our UCCB Bench benchmark dataset consists of 20 culturally grounded categories tailored to reflect Ugandan identity. These include History, Food and Culinary Practices, Ugandan Herbs, Folklore, Value Addition, Streetlife, Slang & Local Expressions, Traditions and Rituals, Values and Social Norms, Attires and Dress Culture, and Festivals. We also include Music, Education, Religion, Architecture, Media, Notable Key Figures, Customs, Literature, and Economy, with each category offering unique insights into Uganda's cultural fabric. The benchmark is going to undergo further expansion with more categories added i.e science , maths, engineering e.t.c.
Figure below provides the full list of UCCB Bench categories.

2.2 Hybrid Data Strategy: Combining Scale with Authenticity (for Fine-tuning)

Addressing the scarcity of large-scale, culturally specific training data (Nekoto et al., 2020; Adebara et al., 2024) requires innovative solutions. We employ a hybrid data strategy for generating our fine-tuning corpus:

2.2.1 Synthetic Data Generation Pipeline

- **Rationale:** To achieve necessary data volume efficiently.
- **Process:** We utilize advanced LLMs (Generator: e.g., Gemma 27B hosted on Afriqloud) prompted with curated Ugandan source texts to generate extensive Question-Answer and instruction-following pairs. Inspired by approaches enhancing synthetic data quality (Zhao et al., 2023; Liu et al., 2024), generated pairs are then evaluated by another AI model (Critiquer: e.g., Command A) based on criteria like factual accuracy (vs. source) and clarity. Only pairs exceeding a quality threshold are retained.

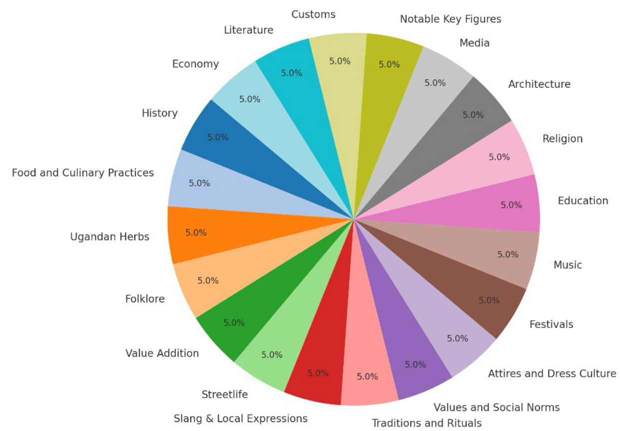

Figure 2: UCCB Bench Categories

2.2.2 Crucial Human Validation

- **Rationale:** AI generation alone cannot guarantee cultural authenticity or capture deep nuance.
- **Process:** A statistically significant sample of the AI-filtered synthetic data undergoes rigorous manual validation by Ugandan elders and cultural experts. This critical step corrects inaccuracies, removes appropriate content, and ensures alignment with local values and lived experience, distinguishing our approach.

2.2.3 Data Governance

All data sourcing and generation adhere to ethical protocols, respecting source permissions and community data rights.

2.3 Efficient Fine-Tuning & Resource Management

To make cutting-edge AI feasible within Uganda's resource landscape, we focus on efficiency:

- **Base Models:** Commencing with Google's Gemma 3 IT family (gemma-3-4b-it, gemma-3-12b-it) (Google, 2024), with plans to adapt future open models from model providers like Cohere (Command A / R) (Cohere , 2024, 2025).
- **PEFT Techniques:** Primarily employing QLoRA (4-bit Quantized Low-Rank Adaptation) (Dettmers et al., 2023) to significantly reduce memory requirements during fine-tuning while preserving performance, enabling training on accessible hardware. We continuously explore advancements in efficient fine-tuning (QuAILoRA, 2024, LoRA Quantization, 2024). Additionally, we are leveraging insights and tools from the InstructLab project (IBM, 2024), an open-source community-driven approach based on IBM Research. InstructLab facilitates the incremental addition of skills and knowledge to various base LLMs, offering advantages in ease of use for teams with diverse expertise levels and potentially reduced computational requirements compared to traditional extensive fine-tuning, aligning strongly with our resource-efficient strategy.

- **Sovereign Infrastructure:** Training is conducted on secure partner infrastructure hosted within Uganda via Afriqloud (e.g., NVIDIA A-series/RTX 4000-class), ensuring data localization and supporting local providers (Afriqloud, 2025).

2.4 Integrated Security Architecture & Responsible Deployment

Ensuring the safety and security of Crane Models is paramount, especially given their deployment within a sovereign context. Our multi-layered approach includes:

- **Input/Output Safeguards:** Implementing mechanisms for prompt filtering and output safety classification using tools inspired by or directly employing frameworks like Gemma Shield (Google LLC, 2025), NVIDIA NeMo (NVIDIA Corporation, 2025), LLaMA Guard (Meta AI, 2023), or capabilities within models like Gemini (Gemini Team, 2023).It is important to recognize that these tools often operate within distinct ecosystems and may not be directly interoperable; our approach involves carefully selecting and integrating specific safeguard components based on compatibility and effectiveness within our deployment stack.

- **Adversarial Testing:** Rigorously evaluating model robustness against misuse through systematic testing using frameworks like Microsoft's PyRIT (Microsoft Corporation, 2024), employing custom prompts tailored to Ugandan cultural and linguistic contexts.

- **Data Access Control & Infrastructure Security:**Utilizing robust authentication (e.g., Keycloak) and secure deployment practices. To ensure fine-grained control over sensitive cultural and linguistic data within the Afriqloud infrastructure, we implement data Access Control via micro-segmentation, leveraging Virtual Application Networks potentially enabled by Red Hat's Skupper project (Red Hat, 2023). This provides data-level security beyond standard infrastructure measures.

- **Expert Guidance:** Our security strategy is informed by ongoing consultation with AI security specialists from Alter Domus, Coinbase, and the broader community.

2.5 Collaborative Ecosystem

Project Crane's success is amplified by a dynamic collaborative ecosystem that integrates essential local knowledge with global technological expertise and best practices. This synergy is crucial for developing robust, culturally relevant, and responsibly deployed AI for Uganda. Our key collaborators and their contributions include:

- **STI Secretariat – Office of the President (Industry 4.0 Bureau) & AI Studio Uganda:** Serves as the initiating body and strategic driver of Project Crane, providing national direction, resources, and coordinating the core Ugandan engineering team focused on model development and innovation.

- **Ugandan Domain Experts & Institutions (e.g., Makerere University, Kyambogo University, National Cultural Centre, etc.):** The foundational pillar providing indispensable cultural knowledge, linguistic expertise, historical context, and validation resources. Experts guide UCCB development, validate synthetic data, and ensure Crane Models authentically reflect Ugandan realities.

- **Google / DeepMind:** Provides access to state-of-the-art foundation models (Gemma family) and associated research insights, forming the technical base for initial Crane Gemma development. Ongoing engagement seeks feedback on adaptation strategies.

- **Cohere:** Offers valuable expertise(Peer reviews from Alejandro Salamanca and Lidiya Murakhovska) in multilingual model development (Aya, Command A, Command R families, etc.), ethical AI frameworks, and open science community engagement, informing our approach to broader language support and responsible AI practices.
- **IBM:** Contributes potential insights into AI governance, Granite Models, methodologies for synthetic data generation (InstructLab concepts), and potentially performance optimization techniques (Kove concepts explored via sovereign cloud partners).
- **RedHat:** Provides deployment expertise (e.g., OpenShift), tools for secure data access control (Skupper), and enables inference acceleration via partner solutions like Neural Magic (Red Hat, 2025).
- **Kove:** Contributes groundbreaking memory virtualization technology, deployed within our partner infrastructure (Afriqloud), enabling significant potential increases in data center performance and power efficiency for demanding AI workloads.
- **Afriqloud:** The crucial partner providing sovereign cloud infrastructure hosted within the region, enabling data localization, local control, and offering regional operational expertise vital for sustainable deployment.
- **Enterprise Neurosystem:** Offers connection to a non-profit open source AI research community, providing invaluable peer review (including reviewers of this document: Dr. Ryan Coffee (Stanford University), Sanjay Aiyagari (Red Hat), David Kypuros (Red Hat), John Overton (Kove), Andy Poling (Kove), Bill Wright (Chair), expertise, and alignment with global open research standards.
- **Weights & Biases:** Contributes best practices and potentially tooling for MLOps, specifically experiment tracking, model versioning, and systematic evaluation during the fine-tuning lifecycle.
- **Hugging Face:** Offers essential platforms for dataset/model hosting, community engagement, access to open-source libraries (transformers, datasets, evaluate), and guidance on benchmark/data documentation standards (e.g., Data Cards).
- **Alter Domus / Coinbase:** Provides critical expertise in AI security best practices, guiding the implementation of our security architecture, adversarial testing protocols, and risk mitigation strategies.(i.e Peer reviews and guidance from Zoumana Cisse)
- **(Optional) UNESCO/Other Global Bodies:** Potential engagement for aligning with global AI ethics standards and understanding international efforts in AI for development and cultural preservation.

This multi-stakeholder collaboration ensures Crane Models are built not only on advanced technology but also grounded in local needs, ethical considerations, and global best practices, fostering a truly Ugandan-led AI initiative with international support and relevance.

3 STRATEGIC APPLICATIONS & ANTICIPATED IMPACT

The development of Crane Models, grounded in Ugandan context and deployed sovereignly, unlocks significant potential across multiple sectors vital to Uganda's development. While our initial focus is on building foundational language and multimodal understanding, the vision extends to specialized applications enabled by current and future capabilities.

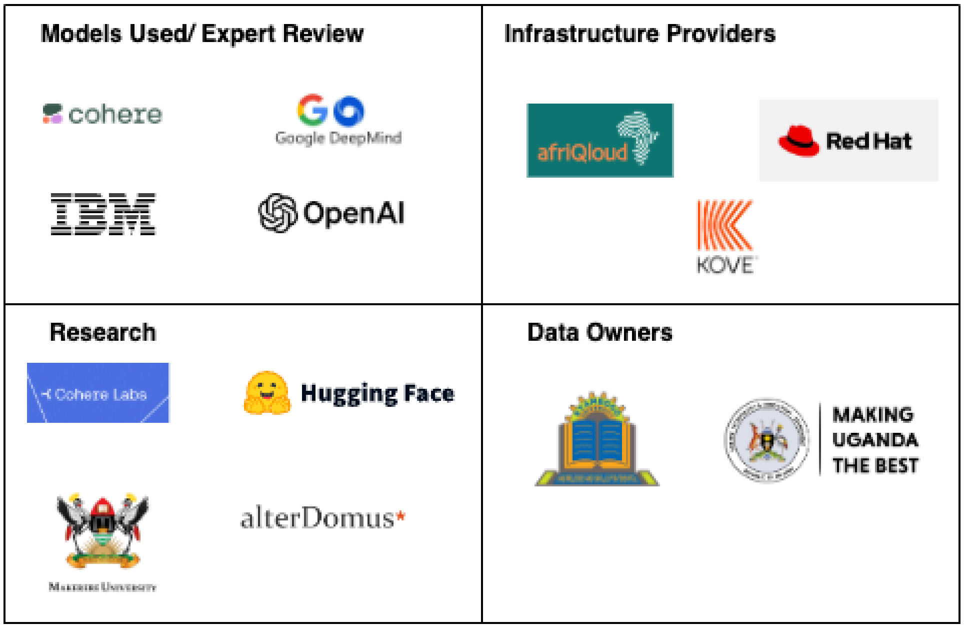

Figure 3: Collaborative Ecosystem Structure

3.1 Educational Applications

Crane Models can directly address gaps left by generic AI in Ugandan education. Immediate applications include:

- **Culturally-Relevant Learning Aids:** Developing interactive tools that accurately explain Ugandan history, lifestyle, civics, literature, and traditional practices, moving beyond potentially biased or incomplete global sources.
- **Language Support:** Assisting learners with nuances of Ugandan English and providing basic support for major local languages (e.g., Luganda, Swahili, Runyakore), improving communication skills.
- **Personalized Content:** Generating explanations or summaries of cultural concepts (proverbs, folklore) tailored to different learning levels (linking to UCCB difficulty tiers).
- **Future Prospect:** Integration into digital learning platforms used by Ugandan schools and universities.

3.2 Agricultural Advisory Services

Agriculture is central to Uganda's economy. Crane Models can provide accessible, relevant information to farmers:

- **Localized Advisory Services:** Offering tailored advice on crop selection, planting schedules, and farming techniques suitable for specific Ugandan regions, soil types, and climates, leveraging knowledge embedded during fine-tuning.
- **Basic Pest/Disease Information:** Providing descriptions and standard mitigation advice for common local agricultural challenges, based on curated agricultural knowledge.
- **Market Linkages:** Offering accessible information on indicative market prices or regional demand for common crops.

3.3 Cultural Heritage Preservation & Access

Crane Models can play a vital role in documenting, preserving, and promoting Uganda's rich cultural heritage:

- **Folklore & Oral Tradition:** Assisting in the transcription, translation, and creation of interactive platforms for exploring traditional stories and oral histories.
- **Language Documentation:** Supporting efforts to document grammar and vocabulary for Uganda's diverse indigenous languages.
- **Virtual Exploration:** Enabling the creation of virtual guides or informational tools for cultural sites and practices.

3.4 Empowering Ugandan Youth & Bridging the Digital Divide

By providing locally relevant AI tools and fostering local AI development capacity, Project Crane aims to:

- **Engage Youth:** Offer technology that resonates with the experiences and language of Uganda's young population (70% under 30), fostering digital literacy and innovation.
- **Enhance Accessibility:** Develop applications potentially deployable on lower-resource devices or with offline capabilities (leveraging efficient models like Gemma 3 QAT variants), improving access for rural communities or those with limited connectivity.
- **Support Linguistic Diversity:** Create foundational technology that can be extended to better serve speakers of Uganda's many languages.

3.5 Future Application Horizons (Prospects & Vision)

While current development focuses on foundational capabilities, the Crane architecture and UCCB framework pave the way for more specialized future applications, potentially leveraging specific model variants or architectures:

- **Environmental Monitoring:** Long-term, specialized models could analyze sensor data for conservation efforts in critical ecosystems like Bwindi Impenetrable Forest (gorilla tracking) and Lake Victoria (water quality). This would likely require integrating or developing dedicated time-series analysis capabilities, potentially inspired by architectures like IBM's Granite Time Series models (IBM, 2024), built upon or alongside the Crane foundation. This remains a prospective goal dependent on further research and data integration.
- **Specialized Domains:** Future Crane Model variants could be fine-tuned for specific professional domains like Ugandan healthcare or legal contexts.
- **Advanced Multilingualism:** Leveraging future, potentially more powerful open models, including anticipated releases from Cohere (building on Command R7B / Command A capabilities) (Cohere AI, 2025) or next-generation Gemma models, to provide deeper support across a wider range of Uganda's indigenous languages.

3.6 Contribution to Sustainable Development Goals (SDGs)

Project Crane aligns directly with several UN SDGs:

- **SDG 4 (Quality Education):** By providing culturally accurate and relevant learning tools.
- **SDG 9 (Industry, Innovation, and Infrastructure):** By building local AI capacity, fostering innovation within Uganda's tech ecosystem, and promoting sovereign digital infrastructure.
- **SDG 10 (Reduced Inequalities):** By making advanced AI technology accessible and relevant to diverse linguistic and socio-economic groups within Uganda, helping to bridge the digital divide.

4 ROADMAP & NEXT STEPS

4.1 Current Status (As of April 2025)

- **UCCB Suite Development:** Core components (UCCB-Knowledge, UCCB-Nuance) designed, with initial data curation and validation using over 25+ Ugandan experts completed. Benchmark structure finalized (Reuel et al., 2024).
- **Synthetic Data Pipeline:** Operational, generating fine-tuning data based on Ugandan sources, with AI critique and crucial elder/expert validation integrated.
- **Initial Model Training:** Crane Gemma 4B fine-tuning, using gemma-3-4b-it as base with QLoRA (Dettmers et al., 2023; Google AI, 2024), is actively underway on the Afriqloud sovereign infrastructure (Afriqloud, 2025). Performance monitored against UCCB Validation set.

4.2 Immediate Next Steps (Near-Term Horizon)

- **Complete v1.0 Fine-tuning & Evaluation:** Finalize the SFT runs for Crane Gemma 4B and 12B. Conduct rigorous evaluation against the full UCCB Test Set v1.0, establishing baseline comparisons against original Gemma models and other relevant open models (e.g., Command R variants, Llama 4 variants, DeepSeek variants, etc.).
- **UCCB Public Release:** Prepare and publicly release UCCB Suite v1.0 components (Development Set, Validation Set, comprehensive Documentation, and Evaluation Scripts) via accessible platforms like Hugging Face Hub (Hugging Face, 2024), `https://huggingface.co/STIAIStudio` to foster community engagement and standardized evaluation.
- **Research Dissemination:** Finalize the internal detailed research paper documenting the UCCB development, synthetic data pipeline, fine-tuning methodology, security framework, and initial Crane Gemma v1.0 results, targeting submission to a top-tier conference (e.g., NeurIPS).
- **Pilot Application Planning:** Initiate detailed planning for deploying Crane Gemma v1.0 in targeted pilot applications in education, healthcare, and agriculture, informed by impact frameworks relevant to African contexts (Google Blog, 2023).

4.3 Future Directions (Mid- to Long-Term Vision)

Our roadmap includes several key expansion vectors:

- **Scaling Crane Models:** Progress to fine-tuning larger parameter models (Gemma 3 27B) as infrastructure and methodology mature. Explore adapting existing or anticipated future open-source releases from companies like Cohere (Cohere AI, 2025), Mistral, and potentially OpenAI, creating diverse Crane Model variants. Explore advanced PEFT techniques like quantization-aware initialization (QuAILoRA, 2024) and information retention methods (LoRA Quantization, 2024).Furthermore, recognizing that foundation model training data (like Gemma's) is not always open, we plan to explore future options for continued pre-training on smaller open base models using our curated Ugandan datasets. This strategy aims to achieve greater model sovereignty and full control over the model's foundational knowledge.
- **Enhanced Reasoning Capabilities:** Investigate fine-tuning techniques or integration with specialized reasoning models to improve performance on the UCCB-Knowledge-HARD subset and complex UCCB-Nuance scenarios. Potentially leveraging efficient kernel implementations (Optimized Kernels Arm CPUs, 2024).
- **Optimized Deployment & Inference:** Leverage quantization (inspired by Gemma 3 QAT (Yvinec & Culliton, 2025)) and deploy Crane Models using highly efficient inference engines suitable for local/sovereign infrastructure, such as Ollama, gemma.cpp (ggerganov, 2024) or Apple's MLX (Apple, 2024) for broader accessibility on diverse hardware.
- **Expanding Modalities - Audio Integration:** Develop Crane Audio capabilities by fine-tuning or integrating open-source Text-to-Speech (TTS) and Speech-to-Text (STT) models like Sesame CSM (SesameAILabs, 2024) on Ugandan languages and accents, enabling voice-based applications.
- **Broader Linguistic Coverage:** Systematically expand UCCB and Crane Model fine-tuning data to include more of Uganda's 40+ languages, employing cross-lingual transfer learning and advanced synthetic data techniques (potentially exploring IBM's InstructLab methodology). Employing cross-lingual transfer learning (AYA Papers, 2024; Adebara et al., 2024) and community-driven data collection.
- **Specialized Domain Models:** Develop Crane Model variants fine-tuned for specific Ugandan sectors like healthcare (understanding local medical terms), legal contexts, or environmental science. For time-series analysis (e.g., Bwindi/Lake Victoria monitoring), explore adapting architectures like IBM's Granite Time Series (IBM, 2024) or leveraging efficient inference scheduling techniques (Task Scheduling, 2024).
- **Deepened Community Engagement:** Establish ongoing feedback loops with Ugandan users, developers, educators, and cultural experts to continuously refine UCCB and Crane Models. Host workshops and training sessions to build local AI capacity. Establish ongoing feedback loops and host workshops (inspired by models like Cohere's Aya, Gemma, Command, Llama community efforts (Cohere AI, 2025; Google Blog, 2024)) and training sessions using accessible tools like Ollama (Ollama, 2024) to build local AI capacity and continuously refine UCCB/Crane Models.

5 CHALLENGES & OPPORTUNITIES FOR INNOVATION

Developing cutting-edge, culturally-grounded AI within a specific regional context like Uganda inherently presents unique challenges, well-documented within the broader AI and development communities. However, we strategically view these challenges not merely as obstacles, but as significant opportunities for pioneering innovation in resource-conscious, culturally-aware, and inclusive AI development. Our approach transforms these constraints into drivers for novel solutions, anchored in recent research and best practices.

5.1 Data Scarcity & Quality Opportunity for Novel Data Strategies

The scarcity of large-scale, high-quality digital datasets for many non-Western languages and cultural contexts is a well-known bottleneck (Nekoto et al., 2020; Adebara et al., 2024). For Uganda specifically, datasets capturing deep cultural nuances are limited. This necessitates innovation beyond standard data collection. Leveraging the strategic support of the STI Secretariat - Office of the President, Project Crane gains privileged access pathways to crucial national data repositories, including facilitating the digitization of unique Ugandan resources held in archives and government bodies. However, access alone isn't sufficient; transforming raw data into AI-ready formats requires novel strategies. Our hybrid data approach addresses this by combining a synthetic data pipeline with essential expert and elder validation. This ensures scalability while prioritizing cultural authenticity. Furthermore, to address data scarcity systematically and foster a broader ecosystem, we are concurrently developing the Uganda Open Data Platform. This platform will serve as a central, accessible repository for curated datasets relevant to Ugandan contexts, facilitating easier data collection, sharing, and utilization for future AI research and development within the country and beyond. This dual approach – utilizing government backing for initial access and building a sustainable open platform – pushes the boundaries of validated synthetic data techniques while creating lasting data infrastructure for Uganda.

5.2 Computational Constraints Opportunity for Resource-Efficient AI

Limited access to large-scale compute infrastructure is a well-recognized challenge in many African regions (UNESCO, 2023). This reality drives innovation in AI efficiency. Our reliance on Parameter-Efficient Fine-Tuning (PEFT) methods, particularly QLoRA which enables fine-tuning large models (e.g., 65B parameters) on single GPUs via 4-bit quantization (Dettmers et al., 2023), is crucial. We explore further optimizations inspired by techniques like quantization-aware adapters (e.g., QDyLoRA (QDyLoRA, 2023)) to minimize memory footprints. Strategic deployment on accessible local partner infrastructure (Afriqloud) maximizes performance within realistic constraints. This aligns with global research on resource-efficient LLMs (Survey Resource-efficient LLMs, 2024).

5.3 Evolving Cultural Dynamics Opportunity for Dynamic & Adaptive AI

Ugandan culture and language are dynamic; norms shift and vocabulary evolves over time (PNAS, 2023). Static benchmarks and models risk obsolescence and performance degradation due to linguistic dataset drift (Linguistic Dataset Drift, 2023). Ensuring the long-term relevance of UCCB and Crane Models requires building systems for continuous learning and adaptation.

This presents an opportunity to develop frameworks incorporating ongoing community input, methods for iterative dataset updates, and models architected for dynamic incorporation of new cultural information, moving beyond static AI snapshots.

5.4 Multilingual Complexity Opportunity for Inclusive Multilingual Models

Effectively supporting Uganda's rich linguistic landscape (over 40 languages) and common phenomena like code-switching (Agarwal et al., 2024) within a unified model framework is a significant challenge. Benchmarks like IrokoBench's AfriMMLU reveal stark performance gaps for LLMs across many African languages (IrokoBench, 2024). This motivates research into inclusive architectures and fine-tuning strategies. We draw inspiration from large-scale multilingual efforts like Cohere's Aya project, particularly the Aya 101 model covering 101 languages (Cohere AI, 2024). Addressing this requires innovative approaches to cross-lingual transfer learning, careful multilingual data balancing, collaboration with local linguistic experts, and developing evaluation metrics sensitive to Uganda's specific linguistic environment.

Figure 4: Continuous Improvement Loop

5.5 AI Talent Shortage & Skill Gaps Opportunity for Local Capacity Building & Upskilling

Developing, deploying, and maintaining sophisticated AI systems like LLMs requires specialized expertise that is currently scarce globally (World Economic Forum, 2023), and particularly so within many African nations, including Uganda (UNESCO, 2021). The required skill set extends beyond general software engineering to include areas like machine learning research, prompt engineering, NLP specific to local languages, PEFT/fine-tuning techniques, MLOps for large models, AI ethics and fairness evaluation, and AI security (UNESCO, 2025a). While Uganda possesses a growing pool of bright, motivated young engineers (as evidenced by the STI AI Studio team), bridging the gap between existing technical education—often rooted in foundational STEM (UNESCO, 2023)—and the specific, rapidly evolving skills needed for cutting-edge LLM work presents a significant challenge, reflecting broader concerns about skill gaps as barriers to technological transformation (World Economic Forum, 2025).

However, this challenge serves as a powerful opportunity to build focused, indigenous AI talent. Project Crane actively addresses this by structuring our talent development roadmap in alignment with pan-African capacity-building initiatives (O'Neill et al., 2024) through:

- **On-the-Job Training:** Providing the core team of young Ugandan engineers with hands-on experience across the LLM lifecycle – from data curation and pipeline development to fine-tuning, evaluation, and secure deployment on platforms like Afriqloud. This aligns with global trends where training workers to utilize AI is a key priority (World Economic Forum, 2023).
- **Knowledge Transfer:** Facilitating direct interaction and mentorship through our collaborative ecosystem (Google, Cohere, IBM, Weights and Biases, Alter Domus, etc.), enabling Ugandan engineers to learn global best practices and harness emerging technologies for sustainable development (UNESCO, 2025b).
- **Curriculum Development Input:** Using insights and practical requirements identified during Project Crane and UCCB development to inform potential updates to Ugandan university curricula and vocational training programs, helping align educational outputs with AI industry needs (UNESCO, 2025a).
- **Community Building:** Planning workshops and the open release of the UCCB Suite and Crane Models to stimulate broader engagement and upskilling within Uganda's tech community, contributing to a collective transformation of learning and skills development (UNESCO, 2025c).

By directly confronting the skills gap through practical application, strategic partnerships, and community engagement, Project Crane not only develops AI models but also cultivates the human capital essential for Uganda's sustainable AI future.

6 CONCLUSION & CALL TO ACTION

Project Crane and the Crane Models family signify Uganda's proactive step into the future of Artificial Intelligence. By building sovereign capacity and prioritizing cultural relevance through innovative, resource-conscious methodologies, we are creating AI designed by Ugandans, for Uganda. Our development of the UCCB Suite provides an essential tool for rigorous evaluation, ensuring accountability and guiding progress.

The initial training of Crane Gemma 4B is underway, marking a tangible milestone. We are committed to transparency and collaboration, sharing our benchmark components and findings openly.

We invite the global AI community, researchers, developers, potential partners, cultural institutions, and Ugandan stakeholders to join us. Engage with the UCCB benchmark, explore collaborations, and help us build an AI ecosystem that is not only intelligent but also equitable, culturally rich, and truly serves the needs of diverse communities worldwide.

UCCB Repository (Forthcoming): `https://huggingface.co/STIAIStudio`

ACKNOWLEDGMENTS

We extend our sincere gratitude to the following individuals and organizations for their invaluable contributions through feedback and expert review during the development of this white paper:

- **Cohere:** Alejandro Salamanca and Lidiya Murakhovska for their insightful feedback and expert guidance, which significantly shaped the strategic direction of this document.
- **Enterprise Neurosystem:** We extend our deepest gratitude to Dr. Ryan Coffee (Senior Scientist, Stanford University), Sanjay Aiyagari (Red Hat, Enterprise Neurosystem), David Kypuros (Red Hat, Enterprise Neurosystem), John Overton (Kove, Enterprise Neurosystem), Andy Poling (Kove, Enterprise Neurosystem), and Bill Wright (Chair, Enterprise Neurosystem) for generously providing their expertise through detailed peer reviews. Their insightful feedback substantially enhanced the rigor, clarity, and alignment of this work with global open research standards, ensuring the document's quality and relevance.
- **IBM:** For their thorough peer reviews and valuable feedback on enterprise AI integration and governance best practices, which were instrumental in refining our proposed framework.
- **Alter Domus:** Zoumana Cisse for his dedicated peer review and specific guidance on AI security best practices, which directly informed the security considerations presented in this document.
- **Coinbase (Anonymous Contribution):** We also acknowledge the valuable insights received anonymously regarding AI security best practices, which further enriched our security framework.
- **Google DeepMind:** Members of the Gemma and Developer Relations teams for their initial engagement, positive feedback on Project Crane's vision, and guidance on leveraging the Gemma ecosystem.

Their collective input has been instrumental in shaping the content, ensuring the quality, and strengthening the relevance of this document. We also extend our heartfelt thanks to the broader Project Crane team and our partners at the STI Secretariat for their foundational support and unwavering commitment to this initiative.

APPENDIX: REFERENCES

This appendix lists all references cited within this white paper, providing additional context and resources for interested readers. While we aim for accessibility, we also maintain a commitment to transparency and accurate attribution of sources.

References

Adebara, I., Adelani, D., Ahia, O. (2024). PRISM: A methodology for auditing biases in large language models [Preprint]. arXiv. https://arxiv.org/abs/2410.18906

Apple. (2024). MLX (Version 0.17.2) [Computer software]. GitHub. https://github.com/ml-explore/mlx

Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://dl.acm.org/doi/10.1145/3442188.3445922

Collaboration on International ICT Policy for East and Southern Africa. (2025, February 14). An artificial intelligence eco-system for Uganda: Policy brief [Policy brief]. CIPESA. https://cipesa.org/wp-content/files/briefs/An$_{A}rtificial_{I}ntelligence_{E}co - System_{f}or_{U}ganda_{policy_{B}rief.pdf}$

Cohere. (2024). Aya Model: An Instruction Finetuned Open-Access Multilingual Language Model [Preprint]. arXiv. https://arxiv.org/abs/2402.07827

Cohere. (2025). Command A: Enterprise-ready large language model [Preprint]. arXiv. https://arxiv.org/abs/2504.00698

DataReportal. (2025). Uganda digital report. https://datareportal.com/reports/digital-2025-uganda

Google LLC. (2024). Synthetic data generation and multi-step RL for reasoning and tool use [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2504.04736

Dettmers, T., Pagnoni, A., Holtzman, A. (2023).

Google. (2024). Gemma: Open models based on Gemini research and technology. https://ai.google.dev/gemma

Google. (2025). Gemma: Family of models. https://ai.google.dev/gemma

Google. (2024). Community efforts. Google Blog. https://blog.google

Google. (2023). Turning AI's opportunity into reality for Africa. Google Blog. https://blog.google

Google. (2024). Gemma QLoRA guide. https://ai.google.dev/gemma

Google . (2025). Gemma shield. https://ai.google.dev

Gemini Team, Google. (2023). Gemini: A family of highly capable multimodal models [Preprint]. arXiv. https://arxiv.org/pdf/2312.11805

ggerganov. (2024).[Computer software]. GitHub. https://github.com/ggerganov/llama.cpp

Globalization Partners International. (2024, July 12). Uganda's culture, language, customs, and traditions [Blog post]. Globalization Partners International. https://www.globalizationpartners.com/2024/07/12/uganda-culture-language-customs-and-traditions

Hendrycks, D., Burns, C., Basart, S. (2020). MMLU: Measuring massive multitask language understanding [Preprint]. arXiv. https://arxiv.org/abs/2009.03300

Huang, Y., Bai, Y., et. al

DeepSeek.

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., . . . Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language ModelsDeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning[Preprint]. arXiv. https://arxiv.org/abs/2501.12948

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., . . . Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language Models [Preprint]. arXiv. https://arxiv.org/pdf/2402.10946

Nadas, M., Diosan, L., & Tomescu, A. (2025, March 18).C-Eval: A Multi-Level Multi-Discipline Chinese Evaluation Suite for Foundation Models [Preprint]. arXiv. https://arxiv.org/abs/2305.08322

DeepSeek.

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language ModelsDeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning[Preprint]. arXiv. https://arxiv.org/abs/2501.12948

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language Models [Preprint]. arXiv. https://arxiv.org/pdf/2402.10946

Nadas, M., Diosan, L., & Tomescu, A. (2025, March 18). Synthetic data generation using large language models: Advances in text and code [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2503.14023

Ogundepo, O., Gwadabe, T. R., Rivera, C. E., Clark, J. H., Ruder, S., Adelani, D. I., Dossou, B. F., Diop, A. A., Sikasote, C., Hacheme, G., Buzaaba, H., Ezeani, I., Mabuya, R., Osei, S., Emezue, C., Kahira, A. N., Muhammad, S. H., Oladipo, A., Owodunni, A. T., … Adhiambo, S. (2023, May 11). *AfriQA: Cross-lingual open-retrieval question answering for African languages* [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2305.06897

Ruibo Liu, Jerry Wei, Fangyu Liu (2024). *Best practices and lessons learned on synthetic data* [Preprint]. arXiv. https://arxiv.org/abs/2404.07503

Wang, Y., Wang, M., Manzoor, M. A., Liu, F., Georgiev, G., Das, R. J., … Nakov, P. (2024, February 4). *Factuality of large language models: A survey* [Preprint]. arXiv. https://arxiv.org/abs/2402.02420

Makerere University. (2025). *Makerere launches AI and data science centre to drive Africa's technological development*. https://news.mak.ac.ug/2025/03/makerere-launches-ai-and-data-science-centre-to-drive-africas-technological-development/

McKinsey and Company. (2023). *AI index: Measuring AI's progress*. https://www.mckinsey.com/capabilities/quantumblack/our-insights

McKinsey and Company. (2023). *The state of AI in 2023: Generative AI's breakout year*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

McKinsey and Company. (2024). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

Meta AI. (2023). *Llama guard: LLM-based Input-Output Safeguard for Human-AI Conversations* [Preprint]. arXiv. https://arxiv.org/pdf/2312.06674

Microsoft Corporation. (2024). *PyRIT* [Computer software]. https://azure.github.io/PyRIT/

Ministry of ICT and National Guidance. (2022). *Vision for A Digital Uganda*: Digitalization for Uganda's Socio-economic Transformation*QLoRA: Efficient Finetuning of Quantized LLMs [Preprint]. arXiv. https://arxiv.org/abs/2305.14314*

Google. (2024). Gemma: Open models based on Gemini research and technology. https://ai.google.dev/gemma

Google. (2025). Gemma: Family of models. https://ai.google.dev/gemma

Google. (2024). Community efforts. Google Blog. https://blog.google

Google. (2023). Turning AI's opportunity into reality for Africa. Google Blog. https://blog.google

Google. (2024). Gemma QLoRA guide. https://ai.google.dev/gemma

Google . (2025). Gemma shield. https://ai.google.dev

Gemini Team, Google. (2023). Gemini: A family of highly capable multimodal models [Preprint]. arXiv. https://arxiv.org/pdf/2312.11805

ggerganov. (2024).[Computer software]. GitHub. https://github.com/ggerganov/llama.cpp

Globalization Partners International. (2024, July 12). Uganda's culture, language, customs, and traditions [Blog post]. Globalization Partners International. https://www.globalizationpartners.com/2024/07/12/uganda-culture-language-customs-and-traditions

Hendrycks, D., Burns, C., Basart, S. (2020). MMLU: Measuring massive multitask language understanding [Preprint]. arXiv. https://arxiv.org/abs/2009.03300

Huang, Y., Bai, Y., et. al

DeepSeek.

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language ModelsDeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning[Preprint]. arXiv. https://arxiv.org/abs/2501.12948

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language Models [Preprint]. arXiv. https://arxiv.org/pdf/2402.10946

Nadas, M., Diosan, L., & Tomescu, A. (2025, March 18).C-Eval: A Multi-Level Multi-Discipline Chinese Evaluation Suite for Foundation Models [Preprint]. arXiv. https://arxiv.org/abs/2305.08322

DeepSeek.

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language ModelsDeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning[Preprint]. arXiv. https://arxiv.org/abs/2501.12948

Hugging Face. (2024). Hugging Face hub quickstart guide. https://huggingface.co/docs/huggingface$_hub/en/quick-start$

IBM. (2024). Granite time series forecasting models.https://www.ibm.com/granite/docs/models/time-series/

IBM. (2025). Granite models.https://www.ibm.com/granite

*Adelani, D. I., Ojo, J., Azime, I. A., Zhuang, J. Y., Alabi, J. O., He, X., Ochieng, M., Hooker, S., Bukula, A., Lee, E. A., Chukwuneke, C., Buzaaba, H., Sibanda, B., Kalipe, G., Mukiibi, J., Kabongo, S., Yuehgoh, F., Setaka, M., Ndolela, L., … Stenetorp, P. (2024, June 5). *IrokoBench: A new benchmark for African languages in the age of large language models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.03368*

Tao Ji, Bin Guo, Yuanbin Wu, Qipeng Guo, Lixing Shen, Zhan Chen, Xipeng Qiu, Qi Zhang, Tao Gui (2025, February 20). Towards economical inference: Enabling DeepSeek's multi-head latent attention in any transformer-based LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.14837

Chen Cecilia Liu, Anna Korhonen, Iryna Gurevych(2024). Cultural learning-based culture adaptation of language models [Preprint]. arXiv. https://arxiv.org/abs/2504.02953

Cheng Li., Mengzhuo Chen et al. CultureLLM: Incorporating Cultural Differences into Large Language Models [Preprint]. arXiv. https://arxiv.org/pdf/2402.10946

Nadas, M., Diosan, L., & Tomescu, A. (2025, March 18). Synthetic data generation using large language models: Advances in text and code [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2503.14023

Ogundepo, O., Gwadabe, T. R., Rivera, C. E., Clark, J. H., Ruder, S., Adelani, D. I., Dossou, B. F., Diop, A. A., Sikasote, C., Hacheme, G., Buzaaba, H., Ezeani, I., Mabuya, R., Osei, S., Emezue, C., Kahira, A. N., Muhammad, S. H., Oladipo, A., Owodunni, A. T., … Adhiambo, S. (2023, May 11). AfriQA: Cross-lingual open-retrieval question answering for African languages [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2305.06897

Ruibo Liu, Jerry Wei, Fangyu Liu (2024). Best practices and lessons learned on synthetic data [Preprint]. arXiv. https://arxiv.org/abs/2404.07503

Wang, Y., Wang, M., Manzoor, M. A., Liu, F., Georgiev, G., Das, R. J., … Nakov, P. (2024, February 4). Factuality of large language models: A survey [Preprint]. arXiv. https://arxiv.org/abs/2402.02420

Makerere University. (2025). Makerere launches AI and data science centre to drive Africa's technological development. https://news.mak.ac.ug/2025/03/makerere-launches-ai-and-data-science-centre-to-drive-africas-technological-development/

McKinsey and Company. (2023). AI index: Measuring AI's progress. https://www.mckinsey.com/capabilities/quantumblack/our-insights

McKinsey and Company. (2023). The state of AI in 2023: Generative AI's breakout year. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

McKinsey and Company. (2024). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

Meta AI. (2023). Llama guard: LLM-based Input-Output Safeguard for Human-AI Conversations [Preprint]. arXiv. https://arxiv.org/pdf/2312.06674

Microsoft Corporation. (2024). PyRIT [Computer software]. https://azure.github.io/PyRIT/

Ministry of ICT and National Guidance. (2022). Vision for A Digital Uganda: Digitalization for Uganda's Socio-economic Transformation. Government of Uganda. https://ict.go.ug/site/documents/digital-vision-uganda-2040.pdf

Muhammad, S. H., Adelani, D.. AfriSenti: Sentiment analysis for African languages. Papers With Code. https://arxiv.org/abs/2302.08956

Nekoto, W., .et.al Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages [Preprint]. arXiv. https://arxiv.org/pdf/2010.02353

Nghia Trung Ngo, Thien Huu Nguyen. (2024). Zero-shot Cross-lingual transfer learning survey [Preprint]. arXiv. https://arxiv.org/abs/2411.08785

Nilepost. (2025). LugandaGPT article. https://nilepost.co.ug/lugandagpt/237940

NVIDIA Corporation. (2025). NVIDIA NeMo guardrails [Computer software]. https://developer.nvidia.com/nemo-guardrails

Organisation for Economic Co-operation and Development. (2024). AI futures project. https://www.oecd.org/en/topics/policy-issues/artificial-intelligence.html

Organisation for Economic Co-operation and Development. (2025). AI policy dashboard – Uganda. https://oecd.ai/en/dashboards/countries/Uganda

Ollama. (2025). Ollama [Computer software]. Ollama. https://ollama.com

O'Neill, J., Adelani, D., Ahia, O., ... Ruder, S. (2024). AI and the future of work in Africa white paper [Preprint]. arXiv. https://arxiv.org/abs/2411.10091

Arm . (2024). Highly Optimized Kernels and Fine-Grained Codebooks for LLM Inference on Arm CPUs [Preprint]. arXiv. https://arxiv.org/pdf/2501.00032

Myung, J., Lee, N., Zhou, Y., Jin, J., Putri, R. A., Antypas, D., Borkakoty, H., Kim, E., Perez-Almendros, C., Ayele, A. A., Gutiérrez-Basulto, V., Ibáñez-García, Y., Lee, H., Muhammad, S. H., Park, K., Rzayev, A. S., White, N., Yimam, S. M., Pilehvar, M. T., Ousidhoum, N., Camacho-Collados, J., Oh, A. (2024). BLEnD: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages [arXiv preprint arXiv:2406.09948]. https://arxiv.org/abs/2406.09948

PNAS. (2018). The dynamics of norm change in the cultural evolution of language https://www.pnas.org/doi/full/10.1073/pnas.1721059115

Minsoo Kim, Kyuhong Shim, Seongmin Park, Wonyong Sung, Jungwook Choi (2023). Quantization-aware adapters. ACL Anthology. https://aclanthology.org/2023.eacl-main.64/

Neal Lawton, Aishwarya Padmakumar, Judith Gaspers, Jack FitzGerald, Anoop Kumar, Greg Ver Steeg, Aram Galstyan (2024). Quantization-aware initialization [Preprint]. arXiv. https://arxiv.org/abs/2410.14713

Reuel, A., Bucknall, B., Casper, S., Fist, T., Soder, L., Aarne, O., Hammond, L., Ibrahim, L., Chan, A., Wills, P., Anderljung, M., Garfinkel, B., Heim, L., Trask, A., Mukobi, G., Schaeffer, R., Baker, M., Hooker, S., … Trager, R. (2025). Open problems in technical AI governance$[Preprint].arXiv.https : //arxiv.org/pdf/2407.14981$

Saurav Pawar, S.M Towhidul Islam Tonmoy, S M Mehedi Zaman, Vinija Jain, Aman Chadha, Amitava Das. (2024). The what, why, and how of context length extension techniques in large language models: A detailed survey [Preprint]. arXiv. https://arxiv.org/abs/2401.07872

SesameAILabs. (2025). Sesame CSM (Version 1.0) [Computer software]. GitHub. https://github.com/SesameAILabs/csm

Shi, Z., Wang, X., Fan, A. (2020). Artificial intelligence for social good: A survey [Preprint]. arXiv. https://arxiv.org/abs/2001.01818

Stanford Human-Centered Artificial Intelligence Initiative. (2024). Artificial intelligence index report 2024. Stanford Human-Centered AI Institute. https://hai.stanford.edu/ai-index

Zhu, H., Wang, Y., Wang, H., Li, X., & Lin, Z. (2024). Beyond efficiency: A systematic survey of resource-efficient large language models. arXiv Preprint, arXiv:2401.00625. https://arxiv.org/abs/2401.00625

Liu, Y., Wang, X., Xu, X., Zhang, J., & Zhou, J. (2024). The effect of scheduling and preemption on the efficiency of LLM inference serving. arXiv Preprint, arXiv:2411.07447. https://arxiv.org/abs/2411.07447

Time. (2023). Emily M. Bender: The 100 most influential people in AI 2023. https://time.com/collection/time100-ai/6308275/emily-m-bender/

United Nations Development Programme. (2023). AI for sustainable development. https://www.undp.org/digital/ai

United Nations Educational, Scientific and Cultural Organization. (2021). UNESCO launches artificial intelligence capacity-building needs assessment survey for Africa region. https://www.unesco.org/en/articles/unesco-launches-artificial-intelligence-capacity-building-needs-assessment-survey-africa-region

United Nations Educational, Scientific and Cultural Organization. (2023). UNESCO's support to AU's continental AI framework and strategy. https://www.unesco.org/en/articles/unescos-support-aus-continental-ai-framework-and-strategy

United Nations Educational, Scientific and Cultural Organization. (2023). STEM lay the way for AI in Africa. https://unesdoc.unesco.org

United Nations Educational, Scientific and Cultural Organization. (2025). Bwindi impenetrable forest. https://whc.unesco.org/en/list/682/

United Nations Educational, Scientific and Cultural Organization. (2025). Transforming learning and skills development in Africa (2nd continental report). https://unesdoc.unesco.org

United Nations Educational, Scientific and Cultural Organization. (2025). Harnessing emerging technologies for sustainable development in Africa. https://unesdoc.unesco.org

United Nations Educational, Scientific and Cultural Organization. (2025). Collective call to action pre-launch event Kampala. https://unesdoc.unesco.org

United Nations Framework Convention on Climate Change. (2023). Uganda farmer adaptation example. https://unfccc.int

Wang, Y., Kordi, Y., Mishra, S. (2022). Self-instruct: Aligning language models with self-generated instructions [Preprint]. arXiv. https://arxiv.org/abs/2212.10560

World Economic Forum. (2023). Future of jobs report 2023. https://www.weforum.org

World Economic Forum. (2023). Press release on AI/big data training priority. https://www.weforum.org

World Economic Forum. (2025). Future of jobs report 2025: Skills outlook. https://www.weforum.org

Xie, K., Babbar, N., Chen, V., & Turura, Y. (2025, March 11). Enhancing multilingual language models for code-switched input data [Preprint]. arXiv. https://arxiv.org/abs/2503.07990

Yufei Huang, Deyi Xiong. A Chinese Bias Benchmark Dataset Curated with Human-AI Collaboration for Large Language Models [Preprint]. arXiv. https://arxiv.org/abs/2306.16244

Yvinec, E., Culliton, P. (2025). Gemma 3 QAT models: Bringing state-of-the-art AI to consumer GPUs. Google Developers Blog. https://developers.googleblog.com/en/gemma-3-quantized-aware-trained-state-of-the-art-ai-to-consumer-gpus

Zawedde, A. (2025). Uganda's AI governance dilemma: Policy or framework to drive innovation? Nile Post. https://nilepost.co.ug/artificial-intelligence/246224/ugandas-ai-governance-dilemma-policy-or-framework-to-drive-innovation-decision-expected-by-2025

Himanshu Gupta, Kevin Scaria, Ujjwala Anantheswaran, Shreyas Verma, Mihir Parmar, Saurabh Arjun Sawant, Chitta Baral, Swaroop Mishra, TarGEN: Targeted data generation with large language models [Preprint]. arXiv. https://arxiv.org/abs/2310.17876