

# Salomon Kabongo, PhD

847-912-8096 | [kabongosalomon@gmail.com](mailto:kabongosalomon@gmail.com) | [linkedin.com/in/salomon-kabongo](https://www.linkedin.com/in/salomon-kabongo) | [github.com/Kabongosalomon](https://github.com/Kabongosalomon)

Normal, IL | US Permanent Resident (Green Card Holder) – No Sponsorship Required

## EDUCATION

---

### Leibniz Universität Hannover

*PhD in Computer Science, Natural Language Processing*

Hannover, Germany

*Nov. 2020 – Nov. 2025*

### African Master's in Machine Intelligence (AMMI)

*Master's in Machine Intelligence*

Accra, Ghana

*Oct. 2019 – Nov. 2020*

### University of the Western Cape

*Master's in Mathematical Sciences*

Cape Town, South Africa

*Aug. 2018 – Jun. 2019*

### Université de Lubumbashi

*Bachelor of Science (Honours) in Mathematics and Computer Science*

Lubumbashi, DRC

*Oct. 2014 – Jul. 2017*

## EXPERIENCE

---

### Lead Software Engineer

*State Farm - Innovation Group*

Feb. 2022 – Present

*Bloomington-Normal, IL*

- Designed automated pre-labeling pipelines using embedding-based retrieval to accelerate data annotation.
- Architected a proprietary document deduplication system utilizing visual similarity and hashing algorithms to identify near-duplicates, significantly streamlining business workflows.
- Led R&D initiatives on Synthetic Media (Deepfake) detection and Video Understanding; benchmarked Visual Language Models (VLMs) against vendor solutions to inform strategic technology procurement.
- Invented novel computer vision applications for the insurance domain, resulting in **4 patent filings (pending approval)**.

### Board Member

*Masakhane Research Foundation*

2021 – Present

*Global*

- Spearheaded the strategic formation of the **Masakhane AI Hub**, defining the 2025–2029 roadmap to build digital public infrastructure for 1 billion+ African language speakers.
- Secured and currently overseeing the execution of **~\$9M USD in research funding** (including \$5M from the Bill & Melinda Gates Foundation and \$4M from IDRC) to democratize AI access.
- Direct high-level collaborations with strategic partners including **Google.org**, **Lacuna Fund**, and **UNESCO**, scaling the community's impact across 50+ African languages.

### Research Assistant

*L3S/Leibniz Information Center for Science & Technology (TIB)*

Nov. 2020 – Nov. 2022

*Hannover, Germany*

- Engineered the core "Leaderboards" feature for the Open Research Knowledge Graph (ORKG), utilizing Knowledge Graphs to automatically track and visualize state-of-the-art (SOTA) progress across scientific publications.
- Collaborated with Hannover Medical School (MHH) on personalized medicine research, applying machine learning techniques to analyze large-scale genetic datasets for predictive healthcare outcomes.
- Conducted research on Scholarly Information Extraction, developing novel NLP pipelines to extract metric data from unstructured text for knowledge graph construction.

SELECTED PUBLICATIONS & PATENTS

---

<b>Systems and Methods for Advanced Duplicate Image Search and Analysis</b>   <i>US Patent</i>	2024
<ul style="list-style-type: none"><li>• US Patent App. 18/652,500, Publication No. US20240411724A1: Invented a system for identifying duplicate documents using vector embeddings and similarity hashing (Assignee: State Farm).</li></ul>	
<b>Bibletts &amp; LiSTra: African Speech Corpora</b>   <i>Interspeech, NeurIPS Workshops</i>	2022
<ul style="list-style-type: none"><li>• Co-authored "Bibletts", a high-fidelity multilingual speech corpus.</li><li>• Developed LiSTra, the first English-to-Lingala speech translation baseline.</li></ul>	
<b>Automated Mining of Leaderboards for Empirical AI Research</b>   <i>ICADL Best Paper Award</i>	2021
<ul style="list-style-type: none"><li>• Proposed a systematic workflow to extract SOTA performance metrics from unstructured scientific text.</li><li>• Published in ICADL and International Journal on Digital Libraries (30+ citations).</li></ul>	
<b>Participatory Research for Low-Resourced Machine Translation</b>   <i>EMNLP Findings</i>	2020
<ul style="list-style-type: none"><li>• Contributor to the Masakhane NLP initiative; paper has over 280+ citations and set the standard for African Language NLP.</li><li>• Developed benchmarks and datasets that are now standard in the field.</li></ul>	

TECHNICAL SKILLS

---

**Languages:** Python, C/C++, SQL, Bash  
**Deep Learning & AI:** PyTorch, TensorFlow, Hugging Face Transformers, LLMs, RAG, LangChain, OpenCV  
**Cloud & MLOps:** AWS (SageMaker, Lambda), Docker, Kubernetes, Google Cloud Vertex AI, Linux, Git  
**Research Areas:** NLP, Computer Vision, Knowledge Graphs, Speech Translation, Generative AI