

# Salomon Kabongo, PhD

847-912-8096 | [kabongosalomon@gmail.com](mailto:kabongosalomon@gmail.com) | [linkedin.com/in/salomon-kabongo](https://linkedin.com/in/salomon-kabongo) | [github.com/Kabongosalomon](https://github.com/Kabongosalomon)  
Normal, IL | US Permanent Resident (Green Card Holder) – No Sponsorship Required

## EDUCATION

---

<b>Leibniz Universität Hannover</b> <i>PhD in Computer Science, Natural Language Processing</i>	Hannover, Germany Nov. 2020 – Nov. 2025
<b>African Master's in Machine Intelligence (AMMI)</b> <i>Master's in Machine Intelligence</i>	Accra, Ghana Oct. 2019 – Nov. 2020
<b>University of the Western Cape</b> <i>Master's in Mathematical Sciences</i>	Cape Town, South Africa Aug. 2018 – Jun. 2019
<b>Université de Lubumbashi</b> <i>Bachelor of Science (Honours) in Mathematics and Computer Science</i>	Lubumbashi, DRC Oct. 2014 – Jul. 2017

## EXPERIENCE

---

<b>Lead Software Engineer</b> <i>State Farm - Innovation Group</i>	Feb. 2022 – Present Bloomington-Normal, IL
<ul style="list-style-type: none"><li>Designed automated pre-labeling pipelines using embedding-based retrieval to accelerate data annotation.</li><li>Architected a proprietary document deduplication system utilizing visual similarity and hashing algorithms to identify near-duplicates, significantly streamlining business workflows.</li><li>Led R&amp;D initiatives on Synthetic Media (Deepfake) detection and Video Understanding; benchmarked Visual Language Models (VLMs) against vendor solutions to inform strategic technology procurement.</li><li>Invented novel computer vision applications for the insurance domain, resulting in <b>4 patent filings (pending approval)</b>.</li></ul>	
<b>Board Member</b> <i>Masakhane Research Foundation</i>	2021 – Present Global
<ul style="list-style-type: none"><li>Spearheaded the strategic formation of the <b>Masakhane AI Hub</b>, defining the 2025–2029 roadmap to build digital public infrastructure for 1 billion+ African language speakers.</li><li>Secured and currently overseeing the execution of <b>~\$9M USD in research funding</b> (including \$5M from the Bill &amp; Melinda Gates Foundation and \$4M from IDRC) to democratize AI access.</li><li>Direct high-level collaborations with strategic partners including <b>Google.org, Lacuna Fund, and UNESCO</b>, scaling the community's impact across 50+ African languages.</li></ul>	
<b>Research Assistant</b> <i>L3S/Leibniz Information Center for Science &amp; Technology (TIB)</i>	Nov. 2020 – Nov. 2022 Hannover, Germany
<ul style="list-style-type: none"><li>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.</li><li>Collaborated with Hannover Medical School (MHH) on personalized medicine research, applying machine learning techniques to analyze large-scale genetic datasets for predictive healthcare outcomes.</li><li>Conducted research on Scholarly Information Extraction, developing novel NLP pipelines to extract metric data from unstructured text for knowledge graph construction.</li></ul>	

## SELECTED PUBLICATIONS & PATENTS

---

<b>Systems and Methods for Advanced Duplicate Image Search and Analysis   US Patent</b>	2024
• 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).	
<b>Bibletts &amp; LiSTra: African Speech Corpora   Interspeech, NeurIPS Workshops</b>	2022
• Co-authored "Bibletts", a high-fidelity multilingual speech corpus.	
• Developed LiSTra, the first English-to-Lingala speech translation baseline.	
<b>Automated Mining of Leaderboards for Empirical AI Research   ICADL Best Paper Award</b>	2021
• Proposed a systematic workflow to extract SOTA performance metrics from unstructured scientific text.	
• Published in ICADL and International Journal on Digital Libraries (30+ citations).	
<b>Participatory Research for Low-Resourced Machine Translation   EMNLP Findings</b>	2020
• Contributor to the Masakhane NLP initiative; paper has over 280+ citations and set the standard for African Language NLP.	
• Developed benchmarks and datasets that are now standard in the field.	

## 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