Nathan Brown

nbrown9@clemson.edu ❖ Seattle, WA
oxxocodes.github.io ❖ github.com/OxxoCodes ❖ linkedin.com/in/nathanmbro ❖ linktr.ee/nathanakaoxxo

RESEARCH & PUBLICATIONS

- **Pula: Training Large Language Models for Setswana (Brown & Marivate, NAACL 2025)**: Led all training and data efforts, developed the first suite of LLMs for the African language Setswana, curated the largest Setswana corpus and first Setswana instruction-tuning dataset through months of manual curation, web scraping, synthetic data generation, and data augmentation. Pula 14B outperforms GPT-4o on translation tasks and achieves SOTA reasoning capabilities for its size.
- **Efficient Transformer Knowledge Distillation: A Performance Review (Brown et al., EMNLP 2023):** Led the research effort, performed distributed training of four compressed encoder transformer models on a PBS cluster, benchmarked compute costs and downstream performance on several tasks, assisted in developing the GONERD dataset.
- **FAST (2023):** Trained a multi-head attention vision architecture, initially proposed in Smith et al. 2023, on a low-resource dataset of FAST exams, improving performance and allowing healthcare administrators to assess the quality of scans.
- **Hospital Reports (2022):** Fine-tuned BERT-style models to perform tasks such as sentiment analysis and named entity recognition on hospital employee reports, providing healthcare administrators with tools to gain further automated insights into employee concerns, needs, and satisfaction.
- **Deriving the Perimeter of Horizontally and Vertically Stretched Regular Shapes (2019):** Developed a novel method to estimate the circumference of an ellipse and Pi by exactly calculating the perimeter of stretched regular polygons.

WORK EXPERIENCE

Microsoft **October 2024 – Present**
Software Engineer, Microsoft AI *Redmond, WA*

- **Technologies:** Python, PyTorch, ONNX, Azure ML, C#
- Led development of training, evaluation, and data-curation pipelines for multiple small language models, with a focus on synthetic data generation, streamlined training, distillation, quantization, and mobile deployment.
- Owned the design and development of an all-in-one internal LLM evaluation tool for SwiftKey, greatly increasing the standardization, modularity, and usability of our Responsible AI (RAI), Quality, and Latency evaluations.
- Developed tailored manual and synthetic data pipelines to further support product evaluations and model training.
- Red-teamed current and forthcoming LLM-powered features, uncovered unseen gaps in our RAI objectives, enhanced prompt quality and effectiveness, and identified and resolved several multilingual instability issues.

CUhackit **August 2023 – May 2024**
Co-Director *Clemson, SC*

- Led the organization of CUhackit: Clemson University's official student hackathon organization, the only hackathon organization in South Carolina recognized by Major League Hacking and host of the largest hackathon in South Carolina.
- Hosted CUhackit, our flagship hackathon for hackers of all years, and HelloWorld, our freshmen-only hackathon event.
- For the first time since 2020, hosted BothoHacks, our hackathon held in Botswana in partnership with Botho University.
- Managed several teams including Design, Partnerships, Public Relations, Hardware, HackerXP, and HackerHelp while working alongside leading industry partners such as Amazon Web Services (AWS), Deloitte, and Raytheon.

Microsoft **May 2023 – August 2023**
Software Engineering Intern, AMPX
Redmond, WA

- **Technologies:** Azure, MongoDB, React, C#, JavaScript, Python
- Using Azure technologies such as Cosmos DB, Web Applications, and Functions, developed a full-stack web application from scratch to enable service developers to easily submit and share key cloud service metrics such as availability, load, and response time with other developers and teams within the organization.
- Developed a web dashboard and serverless REST API, significantly lowering the barrier to entry for sharing service metrics when compared to other offerings, making integration easier for services that are early in development.
- Owned project from beginning to end, leading all development and design with a focus on streamlining integration for teams prioritizing speed and ease-of-use in early-stage projects.
- Collaborated with another team within my organization to determine required features and use cases and to seamlessly integrate their cloud service into my project, demonstrating its usability in real-world development environments.

Giant Oak **September 2022 – December 2023**

Artificial Intelligence Research Partner *Clemson, SC*

- **Technologies:** Python, PyTorch, Hugging Face
- In partnership with Giant Oak, compressed four state-of-the-art efficient attention encoder transformer models via knowledge distillation, increasing model inference speeds by up to 58% while preserving up to roughly 97% of original model performance, demonstrating an effective method of obtaining high performance and low-cost long-context models.
- Assisted in the development of GONERD, a first-of-its-kind long-context Named Entity Recognition (NER) dataset.
- Conducted resource-intensive parallelized model training using the Clemson University Palmetto Cluster, employing a highly distributed approach across multiple compute nodes and up to 56x NVIDIA A100 GPUs.
- As the lead author and alongside three co-authors, wrote a scholarly publication on our process and findings which was accepted to and presented at EMNLP 2023 in Singapore under the industry track.

Clemson University **August 2022 – May 2023**

Security Analyst Intern *Clemson, SC*

- **Technologies**: Splunk, Microsoft Defender XDR, Elastic, Proofpoint, Abnormal, FireEye, Cherwell, Python
- Discovered a vulnerability that would give attackers administrator access to several of Clemson's internal services, exposing a full map of Clemson's networking infrastructure as well as personal information such as name, address, race, sex, and driver's license number on over 12,000 individuals and 72,000 clients, leading to the vulnerability being patched.
- Performed blue team activities by monitoring and responding to security events and notices, identifying incoming security incidents, detecting compromised systems, reading and writing vulnerability reports, and responding to phishing incidents.

Ally Financial **May 2022 – August 2022**

Software Development Intern *Charlotte, NC*

- **Technologies**: TypeScript, JavaScript, React, Storybook, Docker, Node, Git, HTML, CSS
- Led the complete overhaul of an internal web application, transitioning from Ember.js to React for future development.
- Spearheaded the full visual redesign and redevelopment of the site, working to ensure an enhanced user experience.
- Introduced new features with a focus on quality-of-life improvements, modularity, and performance optimizations.

Clemson University **April 2021 – June 2022**

Virtual Reality Research Assistant *Clemson, SC*

- **Technologies:** Unity, C#, IBM Watson TTS/STT, SQL
- Led development of a cross-platform mobile VR Grand Canyon exploration simulation with a focus on a wide range of target devices, heavy performance optimization techniques, and implementation of new features.
- Led development of a hospital staff training simulation; updated deprecated components and dependencies, debugged various pre-existing issues, enhanced user experience.
- Successfully negotiated additional project budgeting for new hardware, improving our capabilities and allowing us to target a much wider range of modern virtual reality devices.

EDUCATION

Clemson University, Honors College **August 2020 – August 2024**

M.S. and B.S. in Computer Science, Data Science and Informatics concentration, Minor in Cybersecurity *Clemson, SC*

- Graduated with B.S. in May 2024 and M.S. in August 2024, Honors, 3.81/4.0 GPA
- Participated in the IBM Watson in the Watt Creative Inquiry where I worked on FAST and Hospital Reports.
- Competed in the 2020 HelloWorld and 2021 CUhackit Hackathons.
- Formula SAE: Served on the chassis division, focusing on developing structural and composite optimizations on vehicle's impact attenuator using Ansys Mechanical, Altair HyperWorks, and SOLIDWORKS.
- Courses include machine learning, deep learning, computational photography, scientific visualization, and data mining.

SKILLS & INTERESTS

- **Skills:** Python, PyTorch, the Hugging Face ecosystem, LLMs, knowledge distillation, multilingual and low-resource NLP, Azure, distributed computing, C#, JavaScript, and C/C++; strong leadership and interpersonal skills; over a decade of self-directed learning in fields such as artificial intelligence, research, software development, and cybersecurity.
- **Interests:** Natural language processing; model and software efficiency; skiing, hiking, camping, and running.

PROJECTS

- **Finch (WIP)**: A curated high-quality subset of English open-source LLM dialogues, scored via LLM-as-a-judge and classifier ranking, giving researchers a deduplicated, diverse, and data-efficient corpus for general training and annealing.
- **Stawberry**: Curating synthetic data to train LLMs on letter-counting tasks (e.g. Rs in "Strawberry"), challenging the common belief this is a tokenizer-related issue and indicating undertraining may play a larger role than initially perceived.
- **Hydrogen:** Portfolio optimization software utilizing historical quantitative analysis of closing stock prices. Retrieves data from Yahoo Finance, optimizes portfolio based on specified tickers. Supports customizable backtesting on historical data.
- **Kano:** An interpreter for the CHIP-8 programming language with full graphics capabilities utilizing C++ and SDL2.

AWARDS/HONORS

- Eagle Scout; NYLT Staff
- 2nd place in 2019 NCCTM State Fair for "Deriving the Perimeter of Horizontally and Vertically Stretched Regular Shapes"
- Clemson National Scholar Finalist