# Jinda Zhang

+1(236)-965-3670 I jindaznb@gmail.com I Website | Github | Linkedin | Scholar | Vancouver, BC


## SELECTED PUBLICATIONS

"ELLMA: LLM-Powered Conversational AI Chat Agents in VRChat with Social Intelligence" (In Submission) Paper
"Building LLM-based AI Agents in Social Virtual Reality" (CHI LBW '24) Paper Demo
Meaningful system-development using the TORGO dataset for dysarthric ASR (Submitted '24) Paper
Enhancing Tree Type Detection in Forest Fire Risk Assessment: Multi-Stage Approach and Colour Encoding with Forest Fire Risk Evaluation Framework for UAV Imagery Arxiv
Error-correction methods for correction of text transcriptions derived from atypical speech (Submitted '24)

## EDUCATION

Master in Computing (NLP), Northeastern University, Canada, Affiliations: SLangLab, Khoury HCI 23 - Now

- Advisors: Aanchan Mohan, Mirjana Prpa, Michal Aibin, Yvonne Coady, Dakuo Wang
- Research Interest: Generative Models, Multimodal Models, AI for X

Master in Statistics, University of Glasgow, UK 22

Bachelor in Mathematics (Theory), Indiana University Bloomington, USA 20

## RESEARCH EXPERIENCE

**Research Intern** - SLangLab, PI: Prof. Aanchan Mohan FA23 - Now

- **Speech and Natural Language Processing (NLP) Research:**
  • Conducted extensive research on Unified-Modal Speech-Text Pre-Training Models e.g. SpeechT5, Wave2Vec2 and grapheme to phoneme and inverse conversion research under guidance of Prof. Aanchan Mohan
- **Voice Collector Software for Atypical Pattern Speech:**
  • Developed and refined robust **Speech Recognition** and **Voice Conversion** tool, utilizing **Flask** and **React.js,** to capture valuable audio data from users with unique speech patterns, promoting accessibility and inclusivity in speech technology
  • Containerized backend functionalities using **Docker**, orchestrated with Docker Compose for easy deployment and management, incorporated Minio for secure storage of audio recordings with **AWS** S3 compatibility.
- **Error Correction Algorithm and Implementation:**
  • Utilized **BART** LLM, and **G2P** framework on Tatoeba data for word and sentence-level **error correction** in phoneme level
  • Finetune Pre-trained **Wav2Vec** for transcribing Torgo atypical speech, and phonmetic **n-gram** language model for error correction using KenLM, integrated with **CTCDecoder** into existing ASR pipeline for enhanced error correction
- **Novel Multimodal Large Language Models Architecture (MLLMs):**
  • Conduct extensive research on Multimodal Large Language Model such as Speech Large Language Models and Vision Large Language Models for Augmentative and alternative communication (AAC) software for individuals with impaired speech
  • Design a novel **dual encoder** with Wav2Phoneme Encoder + WavLM Encoder + LLM architecture
  • Design Cross Modal **Prompting** techniques to include **phoneme** information into exist Multimodal Speech LLM
- **User Interaction and Social Impact:**
  • Enhancing **Automatic Speech Recognition(ASR)** and voice conversion technologies to benefit individuals suffering from motor speech disorders, with a emphasis on dysarthria, empower individuals with varied speech patterns
- **Linux High Performance Computing and DevOps:**
  • Utilized **Linux** High-Performance Computing (**HPC**) environment for intensive model training and data processing tasks.
  • Developed **Bash** scripts to automate **Slurm** training workflows, optimizing computational resource usage and reducing training time. Leveraged Linux skills to manage HPC resources, monitor job progress.

**Research Intern** - Lapis Lab, University of Illinois Urbana-Champaign FA24 - Now

- Developed MedCalc-Bench, a novel dataset designed to evaluate large language models (LLMs) on over 1000 manually reviewed instances across 55 medical calculation tasks.
- Demonstrated the importance of quantitative reasoning for medical diagnosis, revealing gaps in LLM performance related to arithmetic, rule-based reasoning, and entity extraction in clinical settings.

**Research Intern** - PI: Prof. Mirjana Prpa, Dakuo Wang, Yvonne Coady SP24

- **Human-Computer Interaction (HCI) and NLP Research:**
  • Conducted extensive research on integrating **Large Language Models** to create **believable AI Agents** in **Virtual Reality**
  • Worked under the guidance of Prof. Mirjana Prpa, Prof. Dakuo Wang, Prof. Yvonne Coady focusing on the intersection of virtual reality and AI-driven communication
- **Algorithm Development and Implementation:**
  • Developed and utilized multiple algorithms for calculating **relevance and recency**, enhancing the AI avatar's ability to retrieve relevant memories from database, developed efficient dialogue, event, research assistant module
  • Employed **prompt engineering** techniques to generate realistic and contextually appropriate responses from ChatGPT API
- **AI Avatar Functional Capabilities:**
  • Integrated **Computer Vision** and OCR technologies to equip the AI avatar with observational skills
  • Incorporated Deepgram Nova2 API for **Speech-to-Text (STT)** and OpenAI API for **Text-to-Speech (TTS)** communication
  • Developed a memory system using **MongoDB**, allowing the AI avatar to store and recall interactions as memories
- **System Enhancement:**
  • Developed different **AI Agent** Personas, such as Interviewer, Researcher module using **Prompt engineering**
  • Implemented reflection for Agent to reflect on previous observations to generate higher order and abstract thoughts
- **User Interaction and Social Impact:**
  • Facilitated social connections in the virtual realm, particularly aiding individuals with social awkwardness to engage with the AI agent as a means of forming new, meaningful interactions
  • Enabled the AI agent avatar to interact with users through text, voice, body, and facial expressions
- **Large Language Models (LLM) Judge Evaluation:**
  • Implemented evaluation of System efficiency with memory retrieval using GPT-4, Mistral-13b, Llama-7b as **LLM Judgement**

**Research Intern** - Khoury Research Apprenticeship SP24

- **Computer Vision Research:**
  • Integrates Remotely Piloted Aircraft Systems and **Computer Vision** for sustainable management of Canada's natural resources, with a focus on forests, utilized a novel sliding window technique on georeferenced orthogonal maps
- **Algorithm Development and Implementation:**
  • Generated color maps indicating different fire risks after tree types detection, providing a new tool for fire managers to assess and implement prevention strategies.
  • Used SORT **Object Detection** model for identifying tree types, e.g. YOLO, DETR, Faster RCNN, RTMNet, EfficientNet
- **Model Customization:**
  • Integrated **Convolutional Block Attention Module(CBAM)** for attention mechanism adds-on existing **CNN** architecture
  • Explore effect on using different **color representation** (LUV, Lab, linear RGB, log RGB) as input for object detection tasks inspired by Prof. Bruce Maxwell, applied SOTA optimizations such as **random erasing** for enhanced detection algorithms

• Explore effect of multiple **optimizer**, backbone network structure within the color representation of log RGB data as input

**Mentor, Research Capstone** FA24
- **Lead** 3 teams speech and language researchers, worked as a project lead conducted weekly research check-in meeting

**Teaching Staff, Research Capstone** SU24
- **Lead** a team of 40 student researchers and 9 computer vision **research groups**, worked as a project lead, conducted weekly check-in meeting, research paper revisement, and provide weekly **research agenda** and feedbacks for research groups

**Machine Learning Research Capstone** FA23
- Employed a diverse range of advanced machine learning models, such as **Transformer**(**TabNet, FTTransformers**), **CNN**(**UNet**), **Random Forest**, **XGBoost**, **CatBoost, LSTM, Kernel SVM**, and **KNN**, to predict seasonal wildfires occurrence
- Conducted optimization by **random search**, **bayesian optimization**, and **permutation test** for feature importance analysis
- Utilized **Google Cloud Platform (GCP)** for processing training data and **Vertex AI Workbench** for cloud computing

**Masters Thesis -** Advisor: Prof. Craig Anderson et al. SP22
- Modeling progression of world records, designing **statistical models** such as Generalized Linear Model and data visualization algorithm using statistical **R** packages(ggplot, Tidyverse, Shiny) to analyze patterns of genders and events separately

## TECHNICAL SKILLS

Speech Processing, Natural Language Processing, Computer Vision, Machine Learning System, Statistics(A/B Testing), Generative AI, Python, Java, C++(OpenCV), R, Web Development(JavaScript, TypeScript, Node.js, Express.js, React, React Native, Redux.js), MongoDB, SQL, Android, Google Cloud Platform(GCP), AWS, Git, Unix, Linux, Bash, Docker, Accessibility, Software Testing(UnitTest), Data Analysis, Data Mining, Excel, Cloud Computing, High Performance Computing(Slurm), Computer Systems, Object-Oriented Programming, Search Engine Optimization, Distributed Systems, DevOps, MLOps, CI/CD, Project Management

## INDUSTRY EXPERIENCE

**Data Science Intern** - Happy Prime Inc. FA23 - Now
- Working on Intersection of research & development in field of **Machine Learning**, **Speech and Language Processing**, developing Inclusive, robust and performant speech and language technology for Atypical Speech Speakers

**Freelance Software Engineer** (Contract Agency) **-** Udevu.dev SU22
- Designed and developed engaging and user-friendly **3D** web applications using **Three.JS - Client**: Josh Struve (Udevu)
- Ensured web pages were accessible and complied with **A11y** accessibility standards, website **deployed** at Udevu.dev
- Provided consultative support on **search engine optimization**, ensuring the organization's web presence was optimized

**Software Engineer Intern -** RoarPanda Inc. SP22
- Developed and implemented research tools and algorithms in **Python** to analyze 30,000+ malware software gene samples
- Utilized **Python** and **Django** to analyze and display results of malware software genes and collaborated with the research team to test and debug antivirus software modules for improved user experience.

## PROJECTS

**AI Live!** (Web App)
- Developed a full-stack web application for AI, ML enthusiasts to find projects using **Javascript**, **React, Node.js and Express**, implementing features such as user authentication with **Auth0**, dynamic rendering, and integration with external **API**.
- Utilized **Prisma** ORM to interact with a **MySQL** database, designing database tables to store user information and interactions
- Implemented responsive design principles to ensure the application's usability on cross-platform devices, conducting accessibility testing using **Google Lighthouse** and including **accessibility** reports, including **Jest Testing** Framework

**ML:** FireLive!(Vision), Half-UNet(Paper Reproduce), ForestSeg(VIsion, Segmentation), Summarizer(NLP), VoiceCollector(Flask, React, NLP), Diagnosing Bias in Facial Detection Systems (Vision), Music Generation using Character RNN(LSTM)
**Software:** Financial Analyzer(NLP, OCR), FireLook(Android), Around(GoLang, ElasticSearch)

## SERVICE

**Reviewer (Assistant)**, ICASSP 2024
**Teaching Assistant**, Department of CS, Northeastern University
- Research Capstone(FA24), Research Capstone (SU24), Machine Learning (FA23), Object Oriented Programming (SU23)

**Teaching Assistant**, Department of Mathematics, Indiana University Bloomington
- Trigonometry (19)