Nicholas Saroiu

nsaroiu@hotmail.com — Cell: (206) 939-9398

Dear Dr. Annie En-Shiun Lee,

I am writing to express my interest in your URIEL++ research project. As a third-year Computer Science Specialist at the University of Toronto focusing on Artificial Intelligence, I am eager to contribute to research that aligns with my academic goals.

My passion for computer science stems from fascination with how complex programs emerge from elementary instructions. This interest deepened significantly after taking CSC311 (Introduction to Machine Learning), where I discovered the elegance of mathematical models that capture probabilistic distributions in nature. This experience has motivated my goal to pursue advanced studies in machine learning, with CSC494 representing an important step toward formal research experience.

I'm particularly drawn to your work on the URIEL+ project, which significantly expanded upon the original URIEL project. Your approach to enhancing vector representations of languages by expanding the underlying dataset, implementing more sophisticated imputation algorithms for missing values, and enabling dynamic computation of vector distances demonstrates thoughtful evolution of computational linguistic resources.

My interest in URIEL++ is further fueled by my personal bilingual background. As a second-generation immigrant to the United States with Romanian parents, I acquired Romanian through exposure rather than formal instruction. Without explicit teaching of grammatical rules, I developed an intuitive understanding of the language that has sparked my fascination with comparing Romanian to other Romance languages like French and Spanish. This personal experience has given me an appreciation for the typological similarities and differences across language families that URIEL+ captures and URIEL++ aims to expand on.

Recently, I had an informative meeting with one of your project leads, Mason Shipton, who explained the current focuses of the URIEL++ project. I am particularly interested in the feature dimensionality reduction work and how it can be applied to accelerate distance calculations and other computations that utilize language vector representations. Given my background in machine learning, I'm curious about which dimensionality reduction techniques have proven most effective for preserving the linguistic relationships encoded in these vectors, and how the trade-off between computational efficiency and information preservation is being managed.

I believe I am well-suited for this opportunity for three main reasons:

First, I have established a strong foundation in machine learning through my coursework. Beyond CSC311, I am currently enrolled in CSC413 (Neural Networks and Deep Learning), where I am studying various neural network architectures, addressing training and testing challenges, and exploring optimization techniques. These courses have provided me with the theoretical knowledge necessary to engage meaningfully with research in this domain.

Second, I have applied these concepts in practical projects that demonstrate my technical capabilities. In a recent collaborative project, I worked with three colleagues to develop predictive models for student

performance on diagnostic questions. We implemented and compared three distinct approaches—K-Nearest Neighbors, Item-Response Theory, and Autoencoder Neural Networks—analyzing the relative strengths and limitations of each methodology. This experience enhanced my understanding of machine learning implementation and evaluation in real-world applications.

Third, although my formal research experience at the university level is limited, my participation in the International Baccalaureate program during high school provided me with foundational research skills. The program's emphasis on formal academic writing and independent investigation equipped me with the methodological framework to approach research systematically—skills I am eager to develop further under your mentorship.

Thank you for considering my application. I welcome the opportunity to discuss how I might contribute to your research through this course.

Sincerely,
Nicholas Saroiu