

# NGUYEN KHAI NGUYEN

Lewisburg, Pennsylvania

☎ +272-788-0512 ✉ [nkn002@bucknell.edu](mailto:nkn002@bucknell.edu) 🌐 [nkn002](#) [k](#) [khainguyennguyen](#)

## RESEARCH INTERESTS

---

Natural Language Processing and applications of linear programming in deep learning. I am particularly interested in learning the role of multi-modality in language learning, multilingual NLP, and methods of interpreting current language models.

## EDUCATION

---

**Bucknell University**

*Computer Science and Engineering* **GPA: 3.96/4.0**

**Aug 2019 – May 2023**

*Lewisburg, Pennsylvania*

## PUBLICATIONS

---

**Currently Under Review**

**Getting away with more network pruning: From sparsity to geometry and linear regions**

Junyang Cai, **Khai-Nguyen Nguyen**, Aidan Good, Nishant Shrestha, Ruisen Tu, Xin Yu, Thiago Serra

*Submitted to the International Conference on Learning Representations (ICLR) 2023*

**Like a bilingual baby: The advantage of visually grounding a bilingual language model**

**Khai-Nguyen Nguyen**, Zixin Tang, Ankur Mali, M Alex Kelly

*In arxiv 2022*

## PRESENTATIONS

---

**Poster:** “Understanding bilingualism with the multimodal delta-RNN model”

*Kalman Symposium, April 2022*

**Poster:** “Understanding bilingualism with the multimodal delta-RNN model”

*Susquehanna Valley Undergraduate Research Symposium, July 2021*

## RESEARCH EXPERIENCES

---

**Bucknell Analytics Lab**

*Research Assistant*

**Feb 2022 – current**

*Lewisburg, Pennsylvania*

- Co-developed a pruning method based on the linear region’s upper bound, maintaining 80 – 90% accuracy for highly pruned neural networks on MNIST and Fashion-MNIST
- Assisted developing a theorem on the lower upper bound of the number of linear regions of perceptrons and convolutional neural networks with feedforward layers
- Implemented mixed integer linear programming programs counting the number of linear regions of pruned linear neural networks and identifying stable neurons using gurobipy

**Cognitive Modeling and AI Lab**

*Undergraduate Researcher*

**May 2021 – current**

*Lewisburg, Pennsylvania*

- Built a multimodal multilingual LSTM model from scratch in PyTorch using pretrained ResNet for visual grounding and multiBPE embedding for multilingualism
- Evaluated the model’s language generation through perplexity on the test set and its output embedding on the task of word similarity judgment on WordSim-353, MEN, RG-65 and SimLex datasets
- Developed a 200,000-caption Spanish-English image captioning dataset based on the MSCOCO-ES dataset

## PROJECTS

---

**Evaluating Student Writing**

**Feb 2022**

- Fine-tuned Longformer model on 15,600 training essays from Kaggle’s Feedback Prize Competition
- Achieved 0.596 macro F1 score using token classification to identify the essays’ argumentative and rhetorical elements

## IMDB Sentiment Analysis

Jan 2022

- Fine-tuned RoBERTa sequence classification model to a 0.94 accuracy on 25000 training captions and 25000 validation captions from the IMDB sentiment dataset

## Disaster Tweets

Dec 2021

- Fine-tuned BERT sequence classification model on a dataset of 10,000 tweets and tuned the hyperparameters using custom Random Search on 5-Fold cross-validation
- Achieved an accuracy score of 0.83757 (ranked 50/873) in Kaggle's Disaster Tweet as of submission date

## WORK EXPERIENCES

---

### Bucknell's Computer Science Department

Jan 2022 – current

*CSCI 204 Teaching Assistant*

*Lewisburg, Pennsylvania*

- Organize weekly 2-hour help sessions for students taking Algorithms and Data Structures
- Assess performance of 24 students in weekly lab sessions, grade lab assignments, give out comments on improvement and promote critical thinking in problems and approaches

### Bucknell's Teaching and Learning Center

Aug 2021 – May 2022

*Calculus Study Group Leader and Tutor*

*Lewisburg, Pennsylvania*

- Organize a weekly meetings of 90-minute for students in Calculus
- Give strategic advices and facilitate active learning among students to elevate their level of understanding

### VietAbroad Organization

July 2020 — Jan 2021

*Junior Data Specialist*

*Lewisburg, Pennsylvania*

- Collected ranking data from 400 US national universities and liberal arts colleges from online open sources
- Designed the friendliness scores of US schools with 10 team members for the organization's ranking system

## TECHNICAL SKILLS

---

**Languages:** Python, Java, C, SQL, R

**Developer Tools:** VS Code, Jupyter, PyCharm, Anaconda

**Technologies/Frameworks:** Linux, Scrum, Git, Latex, ACT-R, tensorflow, keras, nltk, pyTorch, transformers, scikit-learn, pandas, numpy, seaborn, matplotlib, gurobipy, gensim

## HONORS

---

Tau Beta Pi Honor Society

2021 – present

Bucknell's Program for Undergraduate Research Recipient

2020

Honorable Mention, Mathematical Contest in Modeling

2020

Bucknell University Dean's List

2019 – present

## SERVICES

---

### Co-founder and Secretary of Bucknell's Machine Learning Association

2021 – present

- Connect machine learning enthusiasts at Bucknell by inviting guest speakers from both industry and academia and organizing a machine learning-themed hackathon

### Tresurer of the Vietnamese Student Association

2021 – 2022

- Raised awareness of Vietnamese culture on campus by organizing traditional events such as Mid-Autumn Festival and Lunar New Year Festival
- Communicated with the Bucknell Student Government to manage budget for said events

### Tresurer of the Society for Asian Awareness at Bucknell

2021 – 2022

- Raised awareness of Asian culture on campus and strengthen the Asian community by organizing events such as Asian Gala and Big-and-Little
- Communicated with the Bucknell Student Government to manage budget for said events