

# Khai-Nguyen Nguyen

knguyen07@wm.edu | (+1) 272-788-0512 | Google Scholar | Personal website | LinkedIn

EDUCATION	<b>College of William and Mary</b> <i>Graduate Student in Computer Science</i> • GPA: 3.96/4.00 • Advisor: Prof. Antonio Mastropaolo	Virginia, USA 2023 - current
	<b>Bucknell University</b> <i>B.Sc. in Computer Science and Engineering, minor in Statistics</i> • GPA: 3.94/4.00 ( <i>Summa Cum Laude</i> )	Pennsylvania, USA 2019 - 2023

RESEARCH INTERESTS	Natural Language Processing, Trustworthy AI, Multimodal AI, AI for Healthcare
--------------------	---

SELECTED PUBLICATIONS ♣: equal contribution	<ol style="list-style-type: none"><li>1. <b>Vision Language Models are Biased</b> [pdf] An Vo♣, Khai-Nguyen Nguyen♣, Mohammad Reza Taesiri, Vy Tuong Dang, Anh Totti Nguyen, Daeyoung Kim <i>AI4Math Workshop @ ICML 2025   Submitted to NeurIPS 2025</i></li><li>2. <b>Sentiment Reasoning for Healthcare</b> [pdf] Khai-Nguyen Nguyen♣, Khai Le-Duc♣, Bach Phan Tat, Duy Le, Truong-Son Hy <i>ACL 2025, Industry Track (Oral)</i></li><li>3. <b>Medical Spoken Named Entity Recognition</b> [pdf] Khai Le-Duc, David Thulke, Hung-Phong Tran, Long Vo-Dang, Khai-Nguyen Nguyen, Truong-Son Hy, Ralf Schluter <i>NAACL 2025, Industry Track</i></li><li>4. <b>Real-time Speech Summarization for Medical Conversations</b> [pdf] Khai Le-Duc♣, Khai-Nguyen Nguyen♣, Long Vo-Dang, Truong-Son Hy <i>Interspeech 2024 (Oral)</i></li><li>5. <b>Getting away with network pruning: From sparsity to geometry and linear regions</b> [pdf] Jeffrey Cai♣, Khai-Nguyen Nguyen♣, Nishant Shrestha, Aidan Good, Ruisen Tu, Xin Yu, Shandian Zhe, Thiago Serra <i>CPAIOR 2023   Workshop on Sparsity in Neural Networks, ICLR 2023</i></li><li>6. <b>Resource-Efficient &amp; Effective Code Summarization</b> [pdf] Saima Afrin, Joseph Call, Khai-Nguyen Nguyen, Oscar Chaparro, Antonio Mastropaolo <i>ACM International Conference on AI Foundation Models and Software Engineering (FORGE 2025)</i></li></ol>
--	--

RESEARCH EXPERIENCES	<b>AURA Lab, College of William and Mary, Virginia, USA</b> <i>Research Assistant. Advisor: Prof. Antonio Mastropaolo</i>	2024.10 - current
	<b>Nguyen Lab, Auburn University, Alabama, USA</b> <i>Remote collaborator. Advisor: Prof. Anh Totti Nguyen</i>	2025.01 - 2025.09
	<b>FPT Software AI Center, Vietnam</b> <i>Remote Research Intern. Supervisor: Prof. Truong-Son Hy</i>	2024.05 - 2024.09
	<b>Analytics Lab, Bucknell University, Pennsylvania, USA</b> <i>Undergraduate Research Assistant. Advisor: Prof. Thiago Serra</i>	2022.05 - 2023.01

INDUSTRY EXPERIENCES	<b>Machine Learning Intern, CodaMetrix</b>   Boston, USA 2025.05 - 2025.08 <ul style="list-style-type: none"> <li>Developed a LLM judge to evaluate and refine medical entity annotations, improving performance by 2–4% F1 over human baseline</li> <li>Designed and deployed entity extraction LLMs with DSPy prompt optimization on Databricks, achieving an overall F1 of 70% on Llama-3.1-70B</li> </ul>
	<b>Machine Learning Intern, CodaMetrix</b>   Boston, USA 2024.05 - 2024.08 <ul style="list-style-type: none"> <li>Trained a BERT-based LLM for ICD-10 multilabel classification, outperforming the deployed system by 4% accuracy</li> <li>Accelerate training by 4 times on Databricks using distributed and parallel training</li> </ul>
	<b>Developer, Cazoodle</b>   Illinois, USA 2023.05 - 2023.08 <ul style="list-style-type: none"> <li>Developed a query expansion system that utilize LLMs to <b>comprehend user intent</b> for targeted keyword recommendation</li> <li>Built and deployed a feature-rich Django interface using Elasticsearch BM25 search engine for comprehensive end-to-end testing</li> </ul>
PRESENTATIONS	<b>Sentiment Reasoning for Healthcare</b> <i>Oral presentation at ACL 2025 Industry Track</i>
	<b>Medical Spoken Named Entity Recognition</b> <i>Poster presentation at NAACL 2025 Industry Track</i>
	<b>Real-time Speech Summarization for Medical Conversations</b> <i>Oral presentation at Interspeech 2024 &amp; Poster presentation at MASC-SLL 2024</i>
	<b>Like a bilingual baby: The advantage of visually grounding a bilingual language model</b> <i>Poster presentation at Susquehanna Valley Undergraduate Research Symposium 2022</i>
PROJECTS	<b>Backplane Defect Detection</b> <i>Bucknell University, USA</i> 2022.08 - 2023.05 <ul style="list-style-type: none"> <li>Senior design project; collaborated with industry partners from <i>TE Connectivity</i></li> <li>Finetuned DenseNet and ResNet architectures for defect classification on 1000 training samples with class imbalance, achieving an <b>96.9% accuracy</b> and <b>0.96 F-1</b></li> <li>Applied <b>GradCAM</b> to <b>localize defect</b> and improve model <b>interpretability</b> for users</li> </ul>
AWARDS AND HONORS	<ul style="list-style-type: none"> <li><b>Recipient</b> Computer Science Fellowship, College of William and Mary 2023  <i>Awarded \$500 by the CSCI admissions committee</i></li> <li><b>Dean's List</b> Bucknell University 2019 – 2023  <i>Achived a GPA of 3.6 or higher every semester</i></li> <li><b>Recipient</b>, Program for Undergraduate Research Grant, Bucknell University 2021  <i>Awarded \$4000 for summer research</i></li> <li><b>Honorable Mention</b>, Mathematical Contest in Modeling 2020  <i>Ranked top 25% (3525/13753) among the participants</i></li> </ul>
TEACHING EXPERIENCES	<b>William and Mary:</b> CSCI 140: Intro to Data Science CSCI 304: Computer Organization
	<b>Bucknell University:</b> CSCI 204: Data Structures and Algorithms MATH 201, 202: Calculus I, II