Khanh Nguyen

knguyen@cvc.uab.cat | website | GitHub

EDUCATION

Universitat Autònoma de Barcelona (UAB), Barcelona Oct 2022 – Present

- Ph.D. in Computer Science (expected)
- Supervisor: Prof. Dr. Dimosthenis Karatzas
- Topics: Privacy-preserving AI, Adversarial Robustness, LLMs, VLLMs

Universitat Autònoma de Barcelona (UAB), Barcelona Sep 2020 – Sep 2021

- Master in Computer Vision
- Thesis: *Multi-modal Image Captioning in Wikipedia*. Grade: 9.4/10
- Courses: Optimization Techniques, Machine Learning for Computer Vision, 3D Vision, Visual Recognition, and Video Analysis, among others.

Hanoi University of Science and Technology (HUST), Hanoi Aug 2013 – Aug 2018

- B.Eng. in Electronics and Telecommunications
- Thesis: *Detecting and handling emergency based on UAV-assisted Wireless Sensor Networks*. Grade: 9.0/10

RESEARCH EXPERIENCE

VLR Group – Computer Vision Center, UAB Oct 2021 – Present

Research Interest: Privacy and Safety in Large Language and Vision–Language Models

Multi-modal Image Captioning

- Integrated multi-modal context—text, images, and knowledge graphs—into image captioning pipeline for better image interpretation.
- Develop interpretable image captioning models with multi-modal context using diffusion models. (Ongoing Project)

European Lighthouse on Secure and Safe AI (ELSA)

- Designed and implemented benchmarks for privacy-preserving techniques in vision-language models, focusing on Document Intelligence use cases.
- Applied differential privacy and federated learning to train private document-based models.
- Investigated privacy vulnerabilities in document-based VQA models using membership inference attacks on document data.
- Studying memorization behaviors and designing data extraction attacks on document-based models. (Ongoing Project)
- Exploring jailbreak and privacy attacks targeting LLMs and VLMs. (Ongoing Project)

WORKING EXPERIENCE

Saltlux Inc., Hanoi Jul 2018 – Jul 2020

Data Engineer, Big Data Team

- Designed and developed distributed systems and workbench for automated, real-time collection and indexing of large-scale data from diverse web sources, including rule-based scenarios, RSS feeds, and Open APIs.
- Built data pipelines for efficient querying and processing streaming data to support downstream analysis and training of NLP models.

ACADEMIC SERVICE

1. **Reviewing:** ACM Multimedia 2024, CVPR 2024, AAAI 2025, CVPR 2025, ICCV 2025
2. **Workshop/Competition Organizer**:
   - NeurIPS 2023 – Privacy Preserving Federated Learning Document VQA
   - IEEE SaTML 2025 – Inference Attacks Against Document VQA

PUBLICATIONS

1. **Khanh Nguyen**, Raouf Kerkouche, Mario Fritz, Dimosthenis Karatzas. DocMIA: Document-Level Membership Inference Attacks against DocVQA Models. In *The Thirteenth International Conference on Learning Representations (ICLR) 2025.*

2. **Khanh Nguyen**, Dimosthenis Karatzas. Federated Document Visual Question Answering: A Pilot Study. In *International Conference on Document Analysis and Recognition (ICDAR) 2024.* (Oral)

3. Rubèn Tito*, **Khanh Nguyen***, Marlon Tobaben*, Raouf Kerkouche, Mohamed Ali Souibgui, Kangsoo Jung, Lei Kang, Ernest Valveny, Antti Honkela, Mario Fritz, Dimosthenis Karatzas. Privacy-Aware Document Visual Question Answering. In *International Conference on Document Analysis and Recognition (ICDAR) 2024.*

4. **Khanh Nguyen**, Ali Furkan Biten, Andres Mafla, Lluis Gomez, Dimosthenis Karatzas. Show, Interpret and Tell: Entity-aware Contextualized Image Captioning in Wikipedia. In *Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI) 2023.* (Oral)

TECHNICAL STRENGTHS

Programming Languages	C++, Python, Java
Deep Learning Frameworks	PyTorch, HuggingFace, Accelerate, TensorFlow
Cloud & Deployment	AWS, Docker, Linux-based Systems
Distributed Data & Streaming	Apache Kafka, Apache Storm, Apache Hadoop
Other Tools	Git, Matlab, Latex