

# Anh Dao

[anhdd2000@gmail.com](mailto:anhdd2000@gmail.com) | +358 41 325 1677 | [anh.d.dao@aalto.fi](mailto:anh.d.dao@aalto.fi) | [Portfolio](#)

## Education

---

### M.Sc. in Computer Science

Aalto University, Finland

*August 2024 – Current*

GPA: 4.25 / 5.00

### B.A. in Computer Science

Grinnell College, United States

*August 2019 – May 2023*

GPA: 3.76 / 4.00

**Relevant Coursework:** Information Security, Network Security, Cloud Software and Development, Deep Learning, Machine Learning, Scalable Systems, Biomedical Machine Learning, Linear Optimization, Web Software Development, Probability and Statistics

## Experience

---

### Nokia Solutions & Networks Oy – Finland

*August 2025 – Current*

*R&D Trainee (Node.js/React)*

- Building internal applications in Node.js and React, including front-end development for RAG agents with Vertex AI to support the 4LS team in day-to-day operations
- Developing and integrating LLM-powered tools for internal workflow automation

### Aalto University – Finland

*August 2024 – Current*

*Teaching Assistant (Python/Docker/Kubernetes) – Computer Science Department*

- Deployed online courses through the A+ Learning Management System, developing autograders and exercise content transfer functionalities
- Provided technical and administrative support for 5 Master-level security courses through exercise sessions, grading, and post-grading technical help desks
- Modified the second iteration of CS-E4300 Network Security exercises based on previous year's feedback, utilizing Docker/Kubernetes for containerized lab environments

### Microsec Zrt. – Budapest, Hungary

*June 2022 – August 2022*

*Software Engineer Trainee (Go)*

- Developed an update to the company's PKI certificate reading linter program to properly handle edge cases of subordinate PKI certificates following RFC5280 standards
- Implemented parsing logic for X.509 certificate extensions and validation rules

### Grinnell College – United States

*August 2022 – May 2023*

*Research Assistant (R/Python) – Data Analysis and Social Inquiry Lab*

- Built multivariate statistical models in R using data gathered from games on [Stat2Labs](#), used by over 1000 students in undergraduate statistics courses across 5 higher education institutions
- Helped on average over 150 students and faculty weekly with computing and data-related projects

### Grinnell College – United States

*August 2021 – December 2022*

*Peer Educator (Java/C) – Computer Science Department*

- Hosted weekly mentor sessions for 100+ students across multiple CS courses, providing coding challenges and interactive Q&A sessions on course materials

## Course Projects

---

**Reproducing and Analyzing DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking** (Python/PyTorch)

*September 2025 – December 2025*

- Reproduced and analyzed DiffDock, a diffusion generative model that learns probability distributions over the ligand pose manifold  $\mathcal{M} = R^3 \times SO(3) \times T^m$  for molecular docking
- Evaluated DiffDock achieving 38.2% Top-1 success rate (RMSD  $< 2\text{\AA}$ ) on PDBBind, outperforming AutoDock Vina (26.9%) and deep learning baselines ( $\sim 20\%$ )
- Analyzed DiffDock-L improvements including Confidence Bootstrapping, achieving 50% success rate on PoseBusters benchmark
- Examined mathematical foundations including IGSO(3) distributions for rotational diffusion, wrapped normal distributions for torsion angles, and SE(3)-equivariant neural network architectures
- Investigated the generalization gap on DockGen benchmark revealing performance drop from 38% to 6% on novel protein domains
- *Role: Analysis and manuscript drafting. Collaborators: Quoc-Huy Trinh, Ting Fu*

### **Complex Event Processing with Proper Load Shedding (Python)** *September 2025 – December 2025*

- Implemented state-based load shedding strategy for the OpenCEP framework to handle bursty workloads while maintaining low latency and acceptable recall
- Developed adaptive load shedding using weighted scoring functions considering importance, priority, time criticality, station importance, and chain potential
- Designed state management for partial matches using hash table implementations achieving  $O(1)$  retrieval time complexity
- Implemented hot-path detection patterns using Kleene closure operators on CitiBike dataset ( $\sim 1.8\text{M}$  trips, 10,850 unique bikes)
- Achieved 90-100% recall rates while maintaining latency limits through intelligent partial match pruning
- Developed feedback controller for dynamic drop fraction adjustment based on p90 latency and CPU/memory pressure
- *Role: Semantic/hybrid load-shedding and state-management implementation. Collaborators: Quoc-Huy Trinh, Pablo Rubio*
- *Code: [github.com/huyquoctrinh/Scalablesys-Assignment1](https://github.com/huyquoctrinh/Scalablesys-Assignment1)*

### **Enhance Graph Retrieval-Augmented Generation by Caching Method (Python/Kuzu)** *September 2025 – December 2025*

- Enhanced GraphRAG pipeline by integrating LRU caching mechanism and memory-efficient caching variant combining context pruning and on-the-fly compression
- Implemented Text2Cypher component using dynamic few-shot prompting with TF-IDF-based exemplar selection for semantic similarity matching
- Developed iterative query refinement with database feedback loop for error correction and deterministic post-processing for schema validation
- Achieved highest throughput and lowest latency at scale with memory-efficient LRU caching, demonstrating 60-90% memory savings for large context-heavy entries
- Utilized Kuzu graph database for heterogeneous graph storage with nodes (Laureates, Scholars, Prizes, Institutions) and edges (AFFILIATED\_WITH, AWARDED)
- *Role: Few-shot prompting implementation and dataset creation. Collaborators: Quoc-Huy Trinh, Pablo Rubio*
- *Code: [github.com/huyquoctrinh/ScalableSys-Proj2](https://github.com/huyquoctrinh/ScalableSys-Proj2)*

### **ICS Network Security in the Rise of IIoT Technology (Seminar Paper)** *January 2025 – April 2025*

- Reviewed and analyzed the Purdue Enterprise Reference Architecture (PERA) model for Industrial Control Systems network security and zone-based segmentation
- Examined Zero-trust Architecture (ZTA) principles and implementation challenges in Operational

Technology (OT) environments

- Analyzed network segmentation methods including identity-based segmentation with blockchain, micro-segmentation for east-west traffic control, and Software-Defined Networking (SDN)
- Investigated Manufacturing Usage Descriptions (MUD) for IIoT device security and automated intrusion detection
- Reviewed IT/OT convergence trends and the applicability of machine learning for dynamic access policy enforcement
- *Tutor: Mohit Sethi*

#### **Encrypted File Transfer Protocol** (Python/PyCryptodome)

*March 2022 – April 2022*

- Implemented a secure file messaging protocol using PyCryptodome with Symmetric Key Cryptography following AES-GCM encryption mode
- Designed protocol to ensure message security even against server-side attacks through authenticated encryption

#### **Sudoku Solver** (C/CUDA)

*March 2021*

- Developed an automated Sudoku solving program using CUDA to perform parallel computing on over 100 Sudoku boards simultaneously
- Implemented constraint propagation and backtracking algorithms optimized for GPU execution

#### **Skills**

---

**Languages:** Python, C/C++, Java, Go, JavaScript/TypeScript, R, MATLAB, HTML/CSS, SQL, Cypher

**Frameworks & Libraries:** PyTorch, React, Node.js, Express.js, DSPy, RDKit, scikit-learn, pandas, NumPy

**Tools & Platforms:** Git, Docker, Kubernetes, Google Cloud Platform, Vertex AI, Wireshark, CUDA, PyCryptodome

**Databases:** Kuzu (Graph DB), Neo4j, PostgreSQL

**Concepts:** Machine Learning, Deep Learning, Graph Neural Networks, Diffusion Models, RAG Systems, Complex Event Processing, Network Security, Zero-trust Architecture, PKI/X.509, Cryptography