

ANDRÉ STORHAUG

+47 91 60 14 72 ◇ Trondheim, Norway

PhD candidate, NTNU ◇ andr3.storhaug@gmail.com ◇ linkedin.com/in/andrestorhaug
github.com/andstor ◇ huggingface.co/andstor ◇ scholar.google.com/citations?user=XEhRmSgAAAAJ

EDUCATION

Philosophiae Doctor (PhD) in Computer Science

Norwegian University of Science and Technology (NTNU)

2022 - present

Trondheim, Norway

Advisors: Prof. Jingyue Li, Prof. Zhirong Yang, & Assoc. Prof. Mariusz Nowostawski

Focus: Large language models, Code generation, Security

Master of Science in Engineering (MSE) in Computer Science

Norwegian University of Science and Technology (NTNU)

2020 - 2022

Trondheim, Norway

Thesis: "Secure Smart Contract Code Synthesis with Transformer Models"

Advisor: Prof. Jingyue Li

Bachelor of Engineering (BE) in Computer Science

Norwegian University of Science and Technology (NTNU)

2017 - 2020

Ålesund, Norway

Thesis: "Voxelizer: an Open Source Voxelization Engine"

Advisor: Prof. Ricardo da Silva Torres

EXPERIENCE

PhD candidate

Norwegian University of Science and Technology (NTNU)

Aug 2022 - Aug 2025

Trondheim, Norway

Automatic code generation using large language models to produce high-quality, maintainable, and secure code.

- Developed novel method "vulnerability-constrained decoding" for generating more secure code when using LLMs.
- Trained and evaluated well over 50 LLMs of various sizes, ranging from 350 million to 16 billion parameters.
- Used various training techniques, including fine-tuning, various PEFT methods, and reinforcement learning.
- Efficient multi-GPU training across multiple nodes on HPC cluster with SOTA NVIDIA hardware.
- Agentic systems for code understanding.

Visiting student

CSIRO's Data61

Sep 2024 - Feb 2025

Sydney, New South Wales, Australia

Research visit at CSIRO's Data61, the data and digital specialist arm of Australia's national science agency. Researched agent-based software security patch discovery.

Research Assistant

Norwegian University of Science and Technology (NTNU)

Oct 2017 - Jul 2022

Ålesund, Norway

Part of the [KQMATH](#) Moodle team at NTNU, Department of ICT and Natural Sciences, under the lead of Prof. Hans Georg Schaathun. Involved in the development of several [plugins](#) for the e-learning management system Moodle:

- [JazzQuiz](#): Lets a teacher run a pre-planned quiz, but with the power of improvisation. Used by 512+ sites.
- [ShortMath](#): A Moodle question type for writing beautiful mathematical expressions in WYSIWYG. Used by 282+ sites.
- [CAPQuiz](#): Computer Adaptive Practice quiz system. Used by 60+ sites.

- [QTracker](#): Issue tracking system for Moodle questions. Used by 40+ sites.

Main responsible for ShortMath and QTracker, as well as the [TeX2Max](#) transpilation library for converting LaTeX math to Maxima code. Also implemented continuous integration for all development projects.

E-learning Assistant

Norwegian University of Science and Technology (NTNU)

Sep 2018 - Jun 2022

Trondheim, Norway

Part of small team at NTNU Section for learning support (earlier Centre for Teaching and Learning) providing several critical teaching related services for the 8,000+ employees and 43,000+ students at NTNU:

- Support for most of the digital services and tools offered by NTNU, like Blackboard, Urkund, KASPER, Zoom, Panopto, and Office 365.
- Create and hold physical and digital courses and training sessions for employees and students.
- Create and maintain digital training resources such as wikis and videos.

PUBLICATIONS

Journal articles

- [1] T. Hu, J. Li, B. Li, and A. Storhaug, "Why smart contracts reported as vulnerable were not exploited?" *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 3, pp. 2579–2596, 2025. doi: [10.1109/TDSC.2024.3520554](https://doi.org/10.1109/TDSC.2024.3520554)

Peer-reviewed conference publications

- [1] A. Storhaug, J. Li, and T. Hu, "Efficient avoidance of vulnerabilities in auto-completed smart contract code using vulnerability-constrained decoding," 2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE), Florence, Italy, 2023, pp. 683–693. doi: [10.1109/ISSRE59848.2023.00035](https://doi.org/10.1109/ISSRE59848.2023.00035)
- [2] J. Li, P. H. Meland, J. S. Notland, A. Storhaug, and J. H. Tysse, "Evaluating the impact of ChatGPT on exercises of a software security course," 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), New Orleans, LA, USA, 2023, pp. 1–6. doi: [10.1109/ESEM56168.2023.10304857](https://doi.org/10.1109/ESEM56168.2023.10304857)

Preprints, technical reports, papers under review

- [3] A. Storhaug and J. Li, "Parameter-efficient fine-tuning of large language models for unit test generation: An empirical study," 2024. doi: [10.48550/arXiv.2411.02462](https://doi.org/10.48550/arXiv.2411.02462)

Theses

- [4] A. Storhaug, "Secure smart contract code synthesis with transformer models," Master's thesis, NTNU, 2022. doi: [11250/3015521](https://doi.org/11250/3015521)
- [5] A. Storhaug, "Voxelizer: an open source voxelization engine," Bachelor's thesis, NTNU, 2020. doi: [11250/2663598](https://doi.org/11250/2663598)

ACADEMIC SERVICES

Organizing Committee

- Web Co-Chair, [FSE 2025](#)

Jan 2024 - Jun 2025

Volunteer Work

- Student Volunteer, [FSE 2025](#)

Jun 2025

TECHNICAL SKILLS

Programming & Scripting Languages

Python, JavaScript, TypeScript, Java, SQL, PHP, Swift, Solidity, LaTeX, HTML, CSS

Machine Learning & NLP Frameworks

PyTorch, Transformers, Datasets, Accelerate, DeepSpeed, PEFT, TRL, Smolagents, Numpy, Scikit-learn, NLTK, Pandas, Polars.

Tooling

Git, Docker, Apptainer/Singularity, Slurm Workload Manager, GitHub Actions, Jupyter, ANTLR, Flex/Bison, PEG.js, Eclipse Xtext, TensorBoard, Weights & Biases, Text Generation Inference (TGI), vLLM.

LANGUAGES

Norwegian Bokmål - Native proficiency

Norwegian Nynorsk - Native proficiency

English - Bilingual proficiency

German - Limited working proficiency

PROJECTS

Apps

- **ShiftR**
MacOS menu bar utility that reveals hidden items by shifting overflowed icons, improving usability on notched or small-screen Macs.

Machine learning

- **DeepSpeed Model Memory Calculator**
Tool for calculating theoretical required memory for the various Zero Redundancy Optimizer (ZeRO) configurations.
- **PEFT Unit Test Generation Experiments**
Benchmarking of PEFT methods for unit test generation performance across a range of LLMs. Available: <https://huggingface.co/datasets/andstor/peft-unit-test-generation-experiments>
- **GPT-J 6B Smart Contract Model**
LLM for generating smart contract code for the Ethereum blockchain. doi: [10.57967/hf/1181](https://doi.org/10.57967/hf/1181)
- **GPT-J 6B Smart Contract Vulnerabilities Model**
LLM for generating smart contract code for the Ethereum blockchain with embedded vulnerability detection. doi:[10.5281/zenodo.7595803](https://doi.org/10.5281/zenodo.7595803)
- **Reinforced Code Generation**
Reinforcement learning for code generation using Transformer Reinforcement Learning (TRL).

Datasets

- **Verified Smart Contracts Dataset**
Dataset of real Ethereum smart contracts. doi:[10.57967/hf/1182](https://doi.org/10.57967/hf/1182).
- **Annotated "The Pile" dataset**
Annotated version of the 825 GiB "The Pile" dataset, consisting of 54 programming languages. Available: https://huggingface.co/datasets/andstor/the_pile_github

Parsers, generators, DSLs, programming languages

- **Mush DSL**
Simple, intuitive, and powerful DSL inspired by multiple popular programming languages. Built with Eclipse Xtext, Ecore, and ANTLR.
- **Very Simple Language**
A Very Simple Language (VSL) programming language. Built with Flex and Bison.
- **latex-math-parser**
Complete LaTeX mathematics parser based on PEG parser generator.
- **maxima-codegen**
Maxima Computer Algebra System (CAS) code generator compatible with latex-math-parser AST.
- **solidity-universal-parser**
ANTLR based Solidity parser compatible with all Solidity versions.

GitHub Actions

- **file-existence-action**
GitHub action to check for the existence of files. Used by 44,000+ GitHub repositories.
- **file-reader-action**
GitHub Action to read the contents of a file. Used by 5,300+ GitHub repositories.
- **jsdoc-action**
GitHub Action to build JSDoc documentation. Used by 1,200+ GitHub repositories.
- **copycat-action**
GitHub Action for copying files to other repositories. Used by 800+ GitHub repositories.