Baris Bayazit

baris@cs.toronto.edu | linkedin.com/in/baris-bayazit/ | github.com/bbayazit16 | www.cs.toronto.edu/~baris

Education

University of Toronto — Toronto, ON, Canada
Honours Bachelor of Science in Computer Science. GPA: 3.83 — *September 2023 – May 2027*

Research Experience

Undergraduate Research Project — January 2026 – April 2026
Department of Computer Science, University of Toronto — *Toronto, ON, Canada*

- For-credit supervised project with Prof. Ningning Xie and Maite Kramarz on linear types for metaprogramming.
- Mechanizing definitions and lemmas in the Rocq theorem prover.

Summer Research Intern — May 2025 – August 2025
Department of Computer Science, University of Toronto — *Toronto, ON, Canada*

- Worked full-time under Prof. Xujie Si on three research projects at the intersection of machine learning (LLMs) and programming languages, leading to one accepted paper (LMPL '25 at ICFP) and two ICLR '26 submissions.
- Built VerifyThis-Verus, Rust implementations for various algorithms and challenges, formally proven in Verus.
- Developed benchmarks to evaluate effectiveness of LLMs on software verification tasks in Rocq and developed parsers for Rocq source code to evaluate different ablations.
- Transformed unstructured text such as First-Order Logic and DSLs into Z3 by designing parsers, creating the foundation of RESpecBench, a benchmark for evaluating LLMs on specification generation.

Publications

- **Barış Bayazıt**, Yao Li, and Xujie Si. 2025. A Case Study on the Effectiveness of LLMs in Verification with Proof Assistants. In Proceedings of the 1st ACM SIGPLAN International Workshop on Language Models and Programming Languages (LMPL '25), October 12–18, 2025, Singapore, Singapore. ACM, New York, NY, USA, 15 pages. 10.1145/3759425.3763391.

Under Review

- **B. Bayazıt**, X. Si. "RESpecBench: How reliable is LLM-as-a-judge? Rigorous Evaluation of Specification Generation with Automated Verification." *Under review, ICLR'26.*
- X. Deng, S. C. Zhong, **B. Bayazıt**, A. Veneris, F. Long, X. Si. "VerifyThisBench: Generating Code, Specifications, and Proofs All at Once." *Under review, ICLR'26.*

Research Projects

RESpecBench

- Created the first multi-domain, sound, and automated LLM benchmark with 707 questions and 5 different domains for natural language to specification generation.
- Integrated and developed parsers and sound verifiers using Z3 and Polygon for each domain, ensuring that successful verdicts establish proof of equivalence between the generated and oracle specification.
- Designed experiments to evaluate the effectiveness of LLM-as-a-Judge for specification equivalence.
- Accepted for POPL'26 student research competition and under review for ICLR'26. Link to preprint: https://www.cs.toronto.edu/~baris/respecbench-preprint.pdf

A Case Study on the Effectiveness of LLMs in Verification with Proof Assistants (*LMPL' 25 @ ICFP*)

- First author on LMPL 2025 workshop paper at ICFP, evaluating the effectiveness of various LLMs in the Rocq prover, with Prof. Xujie Si and Prof. Yao Li from Portland State University.
- Constructed dependency and object graphs parsing Rocq source code for hundreds of proofs across two large Rocq projects, Hs-to-Coq and Verdi, running evaluations for proof generation on state-of-the-art LLMs.

VerifyThisBench

- Translated prior VerifyThis competition challenges into Verus, writing specifications, loop invariants, and adapting algorithms to Rust's ownership and borrowing rules, allowing VerifyThisBench to integrate Verus.
- Currently contributing to implement property-based-testing for Verus, providing a more accurate method to evaluate the correctness of LLM-generated specifications.
- Xun Deng, Sicheng Zhong, **Barış Bayazıt**, Andreas Veneris, Fan Long, and Xujie Si. 2025. VerifyThisBench: Generating Code, Specifications, and Proofs All at Once. arXiv preprint arXiv:2505.19271 [cs.SE]. https://arxiv.org/abs/2505.19271

Talks & Presentations

- **A Case Study on the Effectiveness of LLMs in Verification with Proof Assistants**. LMPL Workshop @ *ICFP/SPLASH 2025*, Singapore, Singapore — *Oral presentation, in person.* Oct 2025. DOI: 10.1145/3759425.3763391
- **RESpecBench: How reliable is LLM-as-a-Judge? Rigorous Evaluation of Specification Generation with Automated Verification**. Toronto, Canada. Presented in 3 venues: Department of Computer Science (Aug 2025, Toronto, Canada), ARIA 2025 (Nov 2025, Toronto, Canada), POPL'26 Student Research Competition (Jan 2026, Rennes, France).

Awards & Grants

2026	**CRA Outstanding Undergraduate Researcher Award (HM)**, Computing Research Association	
2025	**SIGPLAN PAC Funding**, ACM SIGPLAN	*$1,500 USD*
2025	**Best Poster Award, Summer Research Showcase**, Department of CS	*$200 CAD*
2025	**Department of Computer Science Research Award**, University of Toronto	*$8,000 CAD*
2024/25	**Dean's List Scholar**, University of Toronto	
2025	**Dr. Roseann Runte Award**, Victoria College	*$1,000 CAD*
2025	**The Bev and Harvey Botting Scholarship**, Victoria College	*$1,000 CAD*
2024	**Salvatore and Marion (Cooper) Brancaccio Scholarship**, Victoria College	*$1,000 CAD*
2023	**Various Hackathons**, ETHGlobal/NewHacks	*$3100 USD*

Service and Community

Service

- **POPL'26 Student Volunteer**, January 2026. Rennes, France.
- **Web Engineering Lead, CSSU**, University of Toronto, 2024–2025. Led University of Toronto Computer Science Student Union web development, helped design orientation activities, and volunteered weekly at the front lounge.

Community & Mentoring

- **SIGPLAN-M Mentee**, ACM SIGPLAN-M Mentoring Program, 2025–2026.
- **PLMW Attendee**, ICFP/SPLASH (2025); POPL (2026).

Technical Skills

Programming Languages: Rust, Python, TypeScript, JavaScript, Java, Solidity, Go, C, SQL, Rocq (Coq)
Frameworks/Libraries: React.js, Next.js, TailwindCSS, Numpy, Keras, Flask, LLVM, Plausible
Developer Tools: Docker, Google Cloud, Firebase, Cloudflare, Jenkins, LaTeX, Git, VSCode, JetBrains IDEs
Languages: English (Fluent), Turkish (Native), German (B1 Goethe-Zertifikat), French (A2)

Software Projects

VeVM *(Rocq)*: A small virtual machine instruction set verified with small step semantics in Rocq Prover.

Cart *(Rust, Inkwell, LLVM)*: A compiler in Rust for a custom language, Cart, featuring a recursive descent parser, type checker and type inference, and an LLVM IR backend.

VicFood *(Python, Typescript, Google Cloud, Docker, Next.js)*: A dining hall menu website for Victoria College used by 150+ residence students daily, with automated scraping and analytics integration.

L2Savings *(Typescript, Next.js, Alchemy API)*: A blockchain analytics tool comparing L1 vs L2 costs.

Mnemo *(Go, Ganache)*: An EVM Paris assembler and toolkit supporting all 143 opcodes, with tools to make chain forking and contract deployment easy.

Hackathons

NewHacks (2023): Web app using JS/TS, Svelte, Firebase. 1st place among 215 participants and 54 projects.

ETHAmsterdam (2022): Awarded 1,500 USD scholarship for flight, placed among 13 winners out of 1100+ attendees and 165 projects. Developed a tool using zero knowledge cryptography for anonymous voting in public blockchains.