Tom Blanchard

tom.blanchard@mail.utoronto.ca

Education

University of Toronto — **Toronto, ON, Canada**
M.Sc in Electrical and Computer Engineering — *Sep. 2024 - ongoing*

- Advised by Prof. Nicolas Papernot
- GPA: 4.0/4.0
- Selected coursework: Algorithm Design Analysis & Theory, Computer Security, Creative Applications of NLP, Generative AI for Images
- Project: GPT4PT, chatbot for finding MBTI personality types

Centrale Lyon — **Lyon, France**
M.Sc in Engineering — *Sep. 2021 - ongoing*

- GPA: 3.4/4.0
- Selected coursework: Machine Learning & Applications, Object-Oriented Programming, Statistics & Probability, Information Science, Fluid Dynamics, Physics of Matter

Lyon 1 University — **Lyon, France**
B.Sc in Mathematics (attended while in Centrale Lyon) — *Sep. 2022 - May 2023*

Lycée Hoche & Lycée Saint-Louis — **Versailles, France**
Preparatory class — *Sep. 2018 - June 2021*

- Three years math and physics preparation for the French Engineering Schools' National Exams

Awards

2025-2027 Edward S. Rogers Scholarship: Electrical and Computer Engineering Department, University of Toronto

Publications

Conference Proceedings

Open LLMs are Necessary for Current Private Adaptations and Outperform their Closed Alternatives: *Vincent Hanke, **Tom Blanchard**, Franziska Boenisch, Iyiola Olatunji, Michael Backes, Adam Dziedzic. Proceedings of the 38th conference Annual Conference on Neural Information Processing Systems.* Code here

Beautiful Images, Toxic Words: Understanding and Addressing Offensive Text in Generated Images: *Aditya Kumar*, **Tom Blanchard***, Adam Dziedzic, Franziska Boenisch. Proceedings of the 40th Annual AAAI Conference on Artificial Intelligence* Code here

CaMeLs Can Use Computers Too: System-level Security for Computer Use Agents: *Hanna Foerster*, **Tom Blanchard***, Kristina Nikolic, Ilia Shumailov, Cheng Zhang, Nicolas Papernot, Yiren Zhao, Robert Mullins. Preprint arXiv 2026*

*Equal Contribution

Experience

ServiceNow — **Montreal, QB, Canada**
Research Scientist (remote) — *Dec. 2024 - March 2025*
Collaborated with the AI Frontier Team in the Microsoft LLM-mail Inject Challenge: a red-teaming competition against an LLM agent for email management, featuring combinations of state-of-the-art defenses. Scored **7/180** teams with combinations of handcrafted and few-shots optimized prompts.

CISPA **Saarbrucken, Germany**
Intern Research Scientist *Feb. 2024 - July 2024*
Advised by Prof. Franziska Boenisch and Prof. Adam Dziedzic

Atos France **Bezons, France**
Intern Applied Scientist *June 2023 - December 2023*
Created a pipeline for extracting entities (names, times etc) and relations from text with a BERT model backbone, aiming at full automation of information extraction of transportation messages for the Paris 2024 Olympic Games. Achieved **> 90% F1-Score** on real data entites and relation extraction.

Programming Skills

Languages: **ML Frameworks:**
Python *PyTorch, Tensorflow*

- Github personal profile here

Talks

"New threats and defenses of AI Agents": Vector Institute & ServiceNow Research

"An approach to secure Computer-Use Agents": Vector Institute