# MATTEO GIOELE COLLU

PHD STUDENT IN BRAIN, MIND AND COMPUTER SCIENCE

## BASIC INFORMATION

Italian
Padova - Italy

### Contacts


- https://collins-115.github.io/webpage/
- matteogioele.collu@phd.unipd.it
- www.linkedin.com/in/matteo-gioele-collu-33b795227


## FIELDS OF INTEREST

- Algorithms and computability
- Machine and Deep Learning
- Security and Privacy of AI
- Mathematics and statistics
- Explainable AI

## PROGRAMMING SKILLS

- Programming languages: **Python, C, C++, Java, OCaml**
- Tools and relevant libraries:
- **TransformerLens, LangChain, Git, Latex, PyTorch, OpenCV, Excel, PySpark**

## LANGUAGES

- English C1/C2 - Certified IELTS Academic
- Italian - Native Speaker
- Spanish B1

## ABOUT ME

I'm focused on the intersection of Explainable AI and Cybersecurity. You see my hunger in exploring the world, you see my foolishness in research and competitions. A disciplined yet unconventional thinker.

## EXPERIENCE

- **CTF - LL_corsairs co-founder**
  - 5th position in LLMail-Inject, Microsoft prompt injection challenge
  - 10th position in LLM-CTF competition SaTML 2024
- **Guest Researcher**
  - **Örebro University (Sweden), 08/2025 - 10/2025**
    Explainability of failure modes in LLMs
- **Guest Researcher**
  - **Radboud University (The Netherlands), 10/2023 - 04/2024**
    Safety of LLMs and Jailbreak Techniques
- **Research Traineeship (Remote)**
  - **University of Aberdeen (Scotland), 02/2021 - 05/2021**
    Development of chatbot for diet coaching

### Publications

- Publish to Perish: Prompt Injection Attacks on LLM-Assisted Peer Review
- Dr. Jekyll and Mr.Hyde: Two Faces of LLMs
- Unaddressed Challenges in Persuasive Dieting Chatbots

### Academic Projects

- Revealing Demographic Bias in LLMs
- Langchain Agent for Cognitive Support in Mild Cognitive Impairment
- Comparison between CNNs in melanoma classification

## EDUCATION

- **PhD student in Brain, Mind and Computer Science**

**Università degli Studi di Padova (Italy), 2024 - now**

**Topic**: Towards Secure Explainable AI and Misuse Prevention in LLMs

- **MSc in Computer Science (110L/110)**

**Università degli Studi di Padova (Italy), 2021 - 2024**

**Topics**: AI & Deep Learning, Security of Machine Learning

**Thesis**: "Dr. Jekyll and Mr. Hyde: Two Faces of LLMs", awarded by *Presidenza del Consiglio dei Ministri* as one of the best cybersecurity thesis in Italy.

- **BSc in Computer Science (110L/110)**

**Università degli Studi di Cagliari (Italy), 2018 - 2021**

- **Cyberchallenge Cybersecurity course**

**Università degli Studi di Cagliari (Italy), 2020**

- Cryptography, Web Security, Network Security and Reverse Engineering