MATTEO GIOELE COLLU

PHD CANDIDATE IN BRAIN, MIND AND COMPUTER SCIENCE

BASIC INFORMATION

21/08/1998
Italian
Cagliari - Italy

Contacts

- +39 391 492 7780
- matteogioele.collu@phd.unipd.it
- www.linkedin.com/in/matteo-gioele-collu-33b795227

Driving Licence

FIELDS OF INTEREST

- Algorithms and computability
- Machine and Deep Learning
- Security and privacy in Machine Learning
- Mathematics and statistics
- Explainable AI

PROGRAMMING SKILLS

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

LANGUAGES

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

ABOUT ME

Experienced Researcher with a keen focus on AI, and its Cybersecurity intersections. My recent endeavors include uncovering vulnerabilities in leading systems like ChatGPT and Gemini, blending analytical rigor with a playful, game-like approach to research. This unique perspective fosters both creativity and out-of-the-box thinking. Nevertheless, I uphold discipline and professionalism as cornerstones of my work ethic, consistently contributing to team synergy and collective success through collaborative efforts.

EXPERIENCE

Guest Researcher

Radboud University, 10/2023 - 04/2024

- Erasmus Traineeship programme
- Safety of Large Language Models, with focus on jailbreaking techniques and alignment
- Use of GPT3.5, GPT4, OpenAI APIs , LLAMA-2, Together APIs, Bard, Gemini-1.5-Flash, and Google AI APIs

Research Traineeship

University of Aberdeen, 02/2021 - 05/2021

- Erasmus Traineeship programme - Philhumans project
- Development of an NLG-based chatbot for diet coaching

Publications

- Dr. Jekyll and Mr.Hyde: Two Faces of LLMs
- Unaddressed Challenges in Persuasive Dieting Chatbots

Academic Projects

- Implementation and comparison between VGG16, ResNet50 and InceptionV3 in melanoma classification
- Custom Meta-heuristic in CPLEX C++ for the TSP
- Python Sudoku Solver with Simulated Annealing

EDUCATION

PhD candidate in Brain, Mind and Computer Science

Università degli Studi di Padova, 2024 - now

Topic: Towards Secure Explainable AI and Misuse Prevention in LLMs

Master Degree in Computer Science (110L/110)

Università degli Studi di Padova, 2021 - 2024

Topics: Artificial Intelligence, Deep Learning, Computer Vision, Optimization, Big Data Engineering, Security of Machine Learning

Thesis: “Dr. Jekyll and Mr. Hyde: Two Faces of LLMs”, shows how personality can direct Large Language Models’ behavior to bypass safety measures.

Bachelor Degree in Computer Science (110L/110)

Università degli Studi di Cagliari, 2018 - 2021

Cyberchallenge Cybersecurity course

Università degli Studi di Cagliari, 2020

- Course with focus on Cryptography, Web Security, Network Security and Reverse Engineering