Emanuele La Malfa

personal page

Work Experience

Post Doctoral Researcher

Jul. 2023 - Current

Benchmarking Large Language Models.

The Alan Turing Institute & University of Oxford, UK

• Principal Investigators: Prof. Michael Wooldridge, Nigel Shadbolt, and Anthony Cohn.

Research Assistant Oct. 2019 – Mar. 2021

Enabling rapid adoption of artificial intelligence through an anonymized data protocol and explainable models.

University of Oxford, UK

- Principal Investigator: Prof. Marta Kwiatkowska.
- Collaboration with GenieAl and funded through the InnovateUK scheme.
- The collaboration led to a paper published at EMNLP 2020.

Education

University of Oxford

Oct. 2019 - Nov. 2023

PhD in Computer Science, supervised by Prof. Marta Kwiatkowska.

Oxford, UK Feb. 2015 – Sept. 2017

Polytechnic University of Milan

Milan, Italy

Master's Degree in Computer Science and Engineering.

Sept. 2011 – Sept. 2014

Polytechnic University of MilanBachelor's Degree in Computer Engineering.

Milan, Italy

Selected Publications

One Language, Many Gaps: Evaluating Dialect Fairness and Robustness of Large Language Models in Reasoning Tasks

ACL 2025 (main track)

Fangru Li, Shaoguang Mao, Emanuele La Malfa, Valentin Hofmann, Adrian de Wynter, Jing Yao, Si-Qing Chen, Michael Wooldridge, Furu Wei

Language Models are Implicitly Continuous

ICLR 2025 (main track)

Samuele Marro, Davide Evangelista, X. Angelo Huang, Emanuele La Malfa, Michele Lombardi, Michael Wooldridge

A Notion of Complexity for Theory of Mind via Discrete World Models

EMNLP 2024 (Findings)

X. Angelo Huang, Emanuele La Malfa, Samuele Marro, Andrea Asperti, Anthony Cohn and Michael Wooldridge

Graph-enhanced Large Language Models in Asynchronous Plan Reasoning

ICML 2024 (main track)

Fangru Lin, Emanuele La Malfa, Valentin Hofmann, Elle Michelle Yang, Anthony Cohn and Janet Pierrehumbert

Deep Neural Networks via Complex Network Theory: a Perspective

IJCAI 2024 (main track)

Emanuele La Malfa, Gabriele La Malfa, Giuseppe Nicosia, Vito Latora

Language-Models-as-a-Service: Overview of a New Paradigm and it its Challenges

Journal of Artificial Intelligence Research (JAIR) - oral presentation at AAAI 2025 - media coverage here and here

Emanuele La Malfa, Aleks Petrov, Frieder Simon, Christoph Weinhuber, Raza Nazar, Anthony Cohn, Nigel Shadbolt and Michael Wooldridge

Language Models Tokenizers Introduce Unfairness Between Languages

NeurIPS 2023 (main track) - website

Aleksandar Petrov, Emanuele La Malfa, Philip Torr, Adel Bibi

The King is Naked: on the Notion of Robustness for Natural Language Processing

AAAI 2022 (main track) - oral presentation -

Emanuele La Malfa, Marta Kwiatkowska

On Guaranteed Optimal Robust Explanations for NLP Models

IJCAI 2021 (main track)

Emanuele La Malfa, Rhiannon Michelmore, Agnieszka Zbrzeny, Nicola Paoletti, Marta Kwiatkowska

Assessing Robustness of Text Classification through Maximal Safe Radius Computation

EMNLP 2020 (Findings)

Emanuele La Malfa, Min Wu, Luca Laurenti, Benjie Wang, Anthony Hartshorn, Marta Kwiatkowska

Pre-prints/Work Under Review

Large Language Models Miss the Multi-Agent Mark

Under review

Emanuele La Malfa, Gabriele La Malfa, Samuele Marro, Jie M. Zhang, Elizabeth Black, Michael Luck, Philip Torr and Michael Wooldridge

Fixed Point Explainability

Under review

Emanuele La Malfa, Jon Vadillo, Marco Molinari and Michael Wooldridge

Understanding the Logical Capabilities of Large Language Models via Out-of-Context Representation Learning

Under review

Jonathan Shaki, **Emanuele La Malfa**, Michael Wooldridge and Sarit Kraus

When Claims Evolve: Evaluating and Enhancing the Robustness of Embedding Models Against Misinformation Edits

Under review

Jabez Magomere, Emanuele La Malfa, Manuel Tonneau, Ashkan Kazemi, Scott Hale

Code Simulation as a Proxy for High-order Tasks in Large Language Models

Under review

Emanuele La Malfa, Christoph Weinhuber, Orazio Torre, Fangru Lin, Angelo X. Huang, Samuele Marro, Anthony Cohn, Nigel Shadbolt and Michael Wooldridge

Jailbreaking Large Language Models in Infinitely Many Ways

Technical report

Oliver Goldstein, Emanuele La Malfa, Felix Drinkall, Samuele Marro, Michael Wooldridge

A Scalable Communication Protocol for Networks of Large Language Models

Under review

Samuele Marro, Emanuele La Malfa, Jesse Wright, Guohao Li, Nigel Shadbolt, Michael Wooldridge, Philip Torr

Grants & Fellowships

Schmidt AI2050 Senior Fellowship

1M \$

Foundations of LLM-based Multi-Agent Systems. PI: Prof. Michael Wooldridge.

Teaching Experience

reaching Experience	
Deep Learning in Healthcare	2024 (HT)
Practical sessions	University of Oxford, UK
Machine Learning	2023 (MT)
Classes	University of Oxford, UK
Probabilistic Model Checking	2023 (MT)
Practical sessions	University of Oxford, UK
Ethical Computing in Practice	2023 (HT)
Practical sessions	University of Oxford, UK
Deep Learning in Healthcare	2023 (HT)
Practical sessions	University of Oxford, UK
Probabilistic Model Checking	2022 (MT)
Practical sessions	University of Oxford, UK
Machine Learning	2021 (MT)
Classes	University of Oxford, UK
Probabilistic Model Checking	2021 (MT)
Practical sessions	University of Oxford, UK
Fundamentals of Computer Science	2016 (OctDec.)
Practical sessions	Polytechnic University of Milan, Italy

Conferences and Workshops Organization

Benchmarking Large Language Models

28/11/2023

Workshop Organizer

The Alan Turing Institute, London, UK

LOD2020, LOD2021, LOD2022, LOD2023, LOD2025

Lake District/Siena

General Chair (2025), Conference Chair (others)

Invited Lectures, Talks, and Presentations

Code Simulation Challenges for Large Language Models

Group Talk

On Robustness for Natural Language Processing

Group Talk

On the Notion of Robustness for Natural Language Processing

Departmental Talk

Robustness for Natural Language Processing

Lecture - Deep Fridays **Explainable AI**

Lecture - Advanced Artificial Intelligence Course

02/02/24

Bocconi, Italy

19/04/2023

ICREA, Barcelona

17/01/2023

King's College University of London, UK

22/04/2022

University of Bologna, Italy

October - December 2023

04/03/2022

April-June 2023

June-May 2022

August 2022

July 2021

2023-2025

2022

Royal Holloway University of London, UK

Academic Reviewing and Volunteering

The Alan Turing Institute - Reviewer

Reviewers for the Turing Fellow Program - Panel "Fundamental Research in Data Science and AI"

Ukrainian Global University - Interviewer

I interviewed Ukrainian students who want to study in a partner university abroad.

The Kharkiv and Przemyśl Project

I spent a week in Przemyśl (Poland) as a volunteer to help refugees who arrived (returned) from (to) Ukraine.

Ukrainian Global University - Interviewer

I interviewed a dozen of prospective undergraduate Ukrainian students who want to study in a partner university abroad.

Eutanasia Legale - Volunteer

I have collected signatures for a referendum to decriminalize euthanasia. The overall campaign gathered 1.2 million valid signatures.

Tutoring and Mentoring

Williams-Exeter Exchange Programme

Tutoring Saad Waheed, Alisa Kanganis, and Simon Socolow (Williams-Exeter Programme exchange students in machine learning).

University of Oxford

Tutoring Edward Kusel and Aleksandar Radoslavov for their part-B projects (undergraduate in Computer Science).

Lead the Future - Mentor 2022-current

Lead the Future helps Italian STEM talents find their path to brilliant careers. I am currently mentoring 8 students.

Mentoring and Students Supervision

Samuele Marro

I co-supervised his Master's thesis (accepted at ICLR'25). Samuele is a PhD student at the Dept. of Engineering, University of Oxford (2024-).

Angelo Huang

I co-supervised his Bachelor's thesis (accepted at EMNLP'24). Angelo is a Master's student in Computer Science at ETH (2025-).

Giovanni Monea

PhD student in Computer Science at Cornell University (2025-).

Lead the Future

Simone Alghisi

PhD student in Information Engineering and Computer Science at the University of Trento.

Lead the Future

Andrea Cerutti

Master's student in Computer Science and Engineering at the Polytechnic University of Milan. Riccardo Inghilleri

Lead the Future

Master's student in Computer Science at the Polytechnic University of Milan.

Orazio Torre

Lead the Future

Bachelor's student in Computer Science at the University of Salerno.

Annalaura Pegoraro Bachelor's student in math at the University of Padova and an exchange student at Berkeley (Summer 2024). Lead the Future

Mattea Busato

Bachelor's student at Bocconi University and an exchange student at the University of Toronto.

Lead the Future Lead the Future

Christoph Weinhuber

Incoming PhD student at the University of Oxford, Dept. of Computer Science (2024-).

Saad Waheed

I supervised Saad for two terms for the Williams-Exeter exchange program (2024) - Saad is a bachelor's student at Williams College (US).

Alisa Kanganis

I supervised Alisa for two terms for the Williams-Exeter exchange program (2024) - Alisa is a bachelor's student at Williams College (US).

Simon Socolow

I supervised Simon for a term for the Williams-Exeter exchange program (2024) - Simon is a bachelor's student at Williams College (US).

Academic Service

Conference Reviewer: ICML, NeurIPS, ICLR, ACL, EMNLP.