

# Emanuele La Malfa

personal page

## Work Experience

### Research Associate

Jul. 2023 – Current

*Benchmarking Large Language Models.*

*Dept. of Computer Science, University of Oxford*

- Principal Investigators: Prof. [Michael Wooldridge](#), [Nigel Shadbolt](#), and [Anthony Cohn](#).
- I conduct research on Large Language Models, with a particular focus on benchmarking their reasoning and planning capabilities.

### 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 [GenieAI](#) 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*

### Polytechnic University of Milan

Feb. 2015 – Sept. 2017

*Master's Degree in Computer Science and Engineering.*

*Milan, Italy*

### Polytechnic University of Milan

Sept. 2011 – Sept. 2014

*Bachelor's Degree in Computer Engineering.*

*Milan, Italy*

## Selected Publications

– 2025 –

### Large Language Models Miss the Multi-Agent Mark

NeurIPS 2025 (position track, acceptance rate 6%)

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

### Language Models are Implicitly Continuous

ICLR 2025 (main track)

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

### 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

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

ACL 2025 (Findings)

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

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

EMNLP 2025 (Findings)

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

– 2024 –

### Language-Models-as-a-Service: Overview of a New Paradigm and 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

### 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

### 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

– 2023-2020 –

### 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

##### Fixed Point Explainability

Under review

**Emanuele La Malfa**, Jon Vardillo, Marco Molinari and Michael Wooldridge

##### 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

##### Artificial Intelligence Safety Fund

450,000 USD

AI Agent Evaluation & Synthetic Content RFP. PIs: **Emanuele La Malfa** and Samuele Marro.

##### Schmidt AI2050 Senior Fellowship

1M USD

Foundations of LLM-based Multi-Agent Systems. PI: Prof. Michael Wooldridge. I wrote part of the grant proposal.

---

#### Teaching 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 (Oct.-Dec.)

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

General Chair (2025), Conference Chair (others)

Lake District/Siena

## Invited Lectures, Talks, and Presentations

<b>Code Simulation Challenges for Large Language Models</b>	02/02/24
<i>Group Talk</i>	Bocconi, Italy
<b>On Robustness for Natural Language Processing</b>	19/04/2023
<i>Group Talk</i>	ICREA, Barcelona
<b>On the Notion of Robustness for Natural Language Processing</b>	17/01/2023
<i>Departmental Talk</i>	King's College University of London, UK
<b>Robustness for Natural Language Processing</b>	22/04/2022
<i>Lecture – Deep Fridays</i>	University of Bologna, Italy
<b>Explainable AI</b>	04/03/2022
<i>Lecture – Advanced Artificial Intelligence Course</i>	Royal Holloway University of London, UK

## Academic Reviewing and Volunteering

<b>The Alan Turing Institute - Reviewer</b>	October - December 2023
<i>Reviewers for the Turing Fellow Program - Panel “Fundamental Research in Data Science and AI”</i>	
<b>Ukrainian Global University - Interviewer</b>	April-June 2023
<i>I interviewed Ukrainian students who want to study in a partner university abroad.</i>	
<b>The Kharkiv and Przemyśl Project</b>	August 2022
<i>I spent a week in Przemyśl (Poland) as a volunteer to help refugees who arrived (returned) from (to) Ukraine.</i>	
<b>Ukrainian Global University - Interviewer</b>	June-May 2022
<i>I interviewed a dozen of prospective undergraduate Ukrainian students who want to study in a partner university abroad.</i>	
<b>Eutanasia Legale - Volunteer</b>	July 2021
<i>I have collected signatures for a referendum to decriminalize euthanasia. The overall campaign gathered 1.2 million valid signatures.</i>	

## Tutoring and Mentoring

<b>Williams-Exeter Exchange Programme</b>	2023-2025
<i>Tutoring Saad Waheed, Alisa Kanganis, and Simon Socolow (Williams-Exeter Programme exchange students in machine learning).</i>	
<b>University of Oxford</b>	2022
<i>Tutoring Edward Kusel and Aleksandar Radoslavov for their part-B projects (undergraduate in Computer Science).</i>	
<b>Lead the Future - Mentor</b>	2022-current
<i>Lead the Future helps Italian STEM talents find their path to brilliant careers. I am currently mentoring 8 students.</i>	

## Mentoring and Student Supervision

<b>Samuele Marro</b>	
<i>I co-supervised his <a href="#">Master's thesis</a>, published at ICLR'25. Samuele is a PhD student at the Dept. of Engineering, University of Oxford.</i>	
<b>Angelo Huang</b>	
<i>I co-supervised his <a href="#">Bachelor's thesis</a>, published at EMNLP'24. Angelo is a Master's student in Computer Science at ETH.</i>	
<b>Ping Zhu</b>	
<i>I supervised his Master's thesis at Oxford. Ping is doing an MSc in advanced computer science at the University of Oxford.</i>	
<b>Alberto Zurini</b>	
<i>I co-supervised, with Alberto Cazzaniga, his Master's thesis. Alberto is a Master's student in Computer Science at the University of Udine.</i>	
<b>Giovanni Monea</b>	
<i>PhD student in Computer Science at Cornell University (2025-).</i>	Lead the Future
<b>Simone Alghisi</b>	
<i>PhD student in Information Engineering and Computer Science at the University of Trento.</i>	Lead the Future
<b>Andrea Cerutti</b>	
<i>Master's student in Computer Science and Engineering at the Polytechnic University of Milan.</i>	Lead the Future
<b>Riccardo Inghilleri</b>	
<i>Master's student in Computer Science at the Polytechnic University of Milan.</i>	Lead the Future
<b>Orazio Torre</b>	
<i>Bachelor's student in Computer Science at the University of Salerno.</i>	Lead the Future
<b>Annalaura Pegoraro</b>	
<i>Bachelor's student in math at the University of Padova and an exchange student at Berkeley (Summer 2024).</i>	Lead the Future
<b>Mattea Busato</b>	
<i>Bachelor's student at Bocconi University and an exchange student at the University of Toronto.</i>	Lead the Future
<b>Christoph Weinhuber</b>	
<i>PhD student at the University of Oxford, Dept. of Computer Science (2024-).</i>	
<b>Saad Waheed</b>	
<i>I supervised Saad for two terms for the Williams-Exeter exchange program (2024) - Saad is a bachelor's student at Williams College (US).</i>	
<b>Alisa Kanganis</b>	
<i>I supervised Alisa for two terms for the Williams-Exeter exchange program (2024) - Alisa is a bachelor's student at Williams College (US).</i>	
<b>Simon Socolow</b>	
<i>I supervised Simon for a term for the Williams-Exeter exchange program (2024) - Simon is a bachelor's student at Williams College (US).</i>	

## Academic Service

---

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

## Other Skills

---

**Programming Languages:** Python, C++. In the past, I used web languages (e.g., Javascript), and low-level languages (C and Assembly x86).

**Libraries, Tools, Frameworks:** PyTorch, CUDA (C++), Numpy, Eigen, Docker.