

# Emanuele La Malfa

personal page

## Work Experience

### Co-director

Sept. 2025 – Current

*Institute for Decentralized AI (IDAI)*

- IDAI is a research institute that develops protocols, tools and methods for decentralized intelligence.
- Grant: [Artificial Intelligence Safety Fund](#) (450,000 USD)

### Research Associate

Jul. 2023 – Current

*Benchmarking Large Language Models.*

*University of Oxford & The Alan Turing Institute*

- 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](#).

## Grants & Fellowships

### Artificial Intelligence Safety Fund

450,000 USD

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

### ARIA: Trust Everything, Everywhere

20,000 USD

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

### Schmidt AI2050 Compute Fund

Compute

*Foundations of LLM-based Multi-Agent Systems. PI: Prof. Michael Wooldridge and **Emanuele La Malfa**. I wrote the grant proposal.*

### Schmidt AI2050 Senior Fellowship

1M USD

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

## Education

### PhD in Computer Science

University of Oxford

*Supervisor: Prof. [Marta Kwiatkowska](#)*

2019 – 2023

### Master's Degree

Polytechnic University of Milan

*Computer Science and Engineering*

2015 – 2017

### Bachelor's Degree

Polytechnic University of Milan

*Computer Engineering*

2011 – 2014

## 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 Micheltore, 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

### Tacit Coordination of Large Language Models

Under review

Ido Aharon, **Emanuele La Malfa**, Michael Wooldridge, Sarit Kraus

### An End-to-end Planning Framework with Agentic LLMs and PDDL

Under review

**Emanuele La Malfa**, Ping Zhu, Samuele Marro, Sara Bernardini, Michael Wooldridge

### Fixed Point Explainability

Under review

**Emanuele La Malfa**, Jon Vadillo, Marco Molinari and 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

Conferences and Workshops Organization

### Benchmarking Large Language Models

Workshop Organizer

**LOD2020, LOD2021, LOD2022, LOD2023, LOD2025**

General Chair (2025), Conference Chair (others)

28/11/2023

The Alan Turing Institute, London, UK

Lake District/Siena/Grosseto

Invited Lectures, Talks, and Presentations

### Code Simulation Challenges for Large Language Models

Group Talk

02/02/24

Bocconi, Italy

### On Robustness for Natural Language Processing

Group Talk

19/04/2023

ICREA, Barcelona

### On the Notion of Robustness for Natural Language Processing

Departmental Talk

17/01/2023

King's College University of London, UK

### Robustness for Natural Language Processing

Lecture – [Deep Fridays](#)

22/04/2022

University of Bologna, Italy

### Explainable AI

04/03/2022

Lecture – Advanced Artificial Intelligence Course

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"

October - December 2023

### Ukrainian Global University - Interviewer

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

April-June 2023

### The Kharkiv and Przemyśl Project

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

August 2022

### Ukrainian Global University - Interviewer

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

June-May 2022

### Eutanasia Legale - Volunteer

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

July 2021

---

## Academic Service

---

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

---

## Teaching Experience

---

### Deep Learning in Healthcare

Practical sessions

2024 (HT)

University of Oxford, UK

### Machine Learning

Classes

2023 (MT)

University of Oxford, UK

### Probabilistic Model Checking

Practical sessions

2023 (MT)

University of Oxford, UK

### Ethical Computing in Practice

Practical sessions

2023 (HT)

University of Oxford, UK

### Deep Learning in Healthcare

Practical sessions

2023 (HT)

University of Oxford, UK

### Probabilistic Model Checking

Practical sessions

2022 (MT)

University of Oxford, UK

### Machine Learning

Classes

2021 (MT)

University of Oxford, UK

### Probabilistic Model Checking

Practical sessions

2021 (MT)

University of Oxford, UK

### Fundamentals of Computer Science

Practical sessions

2016 (Oct.-Dec.)

Polytechnic University of Milan, Italy

---

## Mentoring and Student Supervision

---

### Samuele Marro

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

### Angelo Huang

I co-supervised his [Bachelor's thesis](#), published at EMNLP'24. Angelo is a Master's student in Computer Science at ETH.

### Ping Zhu

I supervised his Master's thesis at Oxford. Ping is doing an MSc in advanced computer science at the University of Oxford.

### Alberto Zurini

I co-supervised, with Alberto Cazzaniga, his Master's thesis. Alberto is a Master's student in Computer Science at the University of Udine.

### Giovanni Monea

PhD student in Computer Science at Cornell University.

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.

Lead the Future

### Riccardo Inghilleri

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

Lead the Future

### Orazio Torre

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

Lead the Future

### Annalaura Pegoraro

Bachelor's student in math at the University of Padova.

Lead the Future

### Mattea Busato

Bachelor's student at Bocconi University.

Lead the Future

### Alessandro Soccol

Master's student at the University of Cagliari.

Lead the Future

**Nathan Gebreamlak**

*I supervised Nathan for a term for the Williams-Exeter exchange program (2025) - Nathan is a bachelor's student at Williams College (US).*

**Charlie Tharas**

*I supervised Charlie for a term for the Williams-Exeter exchange program (2025) - Charlie is a bachelor's student at Williams College (US).*

**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).*

---

**Tutoring and Mentoring**

---

**Williams-Exeter Exchange Programme****2023-2025**

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

**University of Oxford****2022**

*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 9 students.*