

Emanuele La Malfa

personal page

Work Experience

Co-director	Sept. 2025 – Current
<i>Institute for Decentralized AI (IDAI)</i>	
<ul style="list-style-type: none">• 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

<i>Benchmarking Large Language Models.</i>	<i>University of Oxford & The Alan Turing Institute</i>
<ul style="list-style-type: none">• 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
<i>Enabling rapid adoption of artificial intelligence through an anonymized data protocol and explainable models.</i>	<i>University of Oxford, UK</i>

Grants & Fellowships	
Artificial Intelligence Safety Fund	450,000 USD
<i>AI Agent Evaluation & Synthetic Content RFP. PIs: Emanuele La Malfa and Samuele Marro.</i>	
ARIA: Trust Everything, Everywhere	20,000 USD
<i>AI Agent Evaluation & Synthetic Content RFP. PIs: Emanuele La Malfa, Samuele Marro and Angelo X. Huang.</i>	
Schmidt AI2050 Compute Fund	Compute
<i>Foundations of LLM-based Multi-Agent Systems. PI: Prof. Michael Wooldridge and Emanuele La Malfa. I wrote the grant proposal.</i>	
Schmidt AI2050 Senior Fellowship	1M USD
<i>Foundations of LLM-based Multi-Agent Systems. PI: Prof. Michael Wooldridge. I wrote part of the grant proposal.</i>	

Education

PhD in Computer Science	University of Oxford
<i>Supervisor: Prof. Marta Kwiatkowska</i>	<i>2019 – 2023</i>
Master's Degree	Polytechnic University of Milan
<i>Computer Science and Engineering</i>	<i>2015 – 2017</i>
Bachelor's Degree	Polytechnic University of Milan
<i>Computer Engineering</i>	<i>2011 – 2014</i>

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 Aspert, 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

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

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/Grosseto

Invited Lectures, Talks, and Presentations

Code Simulation Challenges for Large Language Models

02/02/24

Group Talk

Bocconi, Italy

On Robustness for Natural Language Processing

19/04/2023

Group Talk

ICREA, Barcelona

On the Notion of Robustness for Natural Language Processing

17/01/2023

Departmental Talk

King's College University of London, UK

Robustness for Natural Language Processing

22/04/2022

Lecture – *Deep Fridays*

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

Academic Service

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

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

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.