

# FAIR – Fingerprinting of LLMs with cognitive models

## Background

Large Language Models (LLMs) are revolutionising on the one hand the way we interact with cyber systems, and on the other hand how we acquire and spread information in the cyber-physical converged world. While this will present unprecedented advantages and opportunities, it will also expose users to innumerable threats. Ways to characterise and, ideally, distinguish between content generated by LLMs and real humans are therefore fundamental to control such extremely powerful technology.

## Topic description

Fingerprinting of LLMs would be a fundamental tool to tell apart whether content is generated by human beings or not. Clearly, many ways are being explored to fingerprint LLMs. The objective of this topic is to exploit models of the way humans allocate cognitive efforts in the use of words, and analyse whether LLMs produce similar fingerprints, or not. The activities will build on substantial experience of the supervisors on characterisation of human behaviour in large-scale online platforms, such as OSNs. The starting point of this activity will be the content generated by ChatGPT but other LLMs will be also taken into account.

We are looking for a graduate research fellow exploring this topic. The activities will involve a mix of data collection, identification of suitable large-scale datasets, development of data analytics tools, analysis of the obtained results and definition of fingerprints of LLMs outputs.

## Type of prospect positions

We plan to open positions at the level of Graduate Research Fellow on this topic.

## Funding and partnerships

*Supported by:* [FAIR](#), Extended Partnership on Artificial Intelligence (funded by the National Recovery and Resilience Plan (NRRP), European Union - NextGenerationEU)

## Candidate profile

Ideal candidates should have or about to obtain a PhD in Computer Science, Computer Engineering, Mathematics, Physics, or closely related disciplines, and a proven track record of excellent University grades. Preferably, the MSc should be in one of the relevant research areas: Artificial Intelligence, BigData analytics.

## Contacts



**Andrea Passarella**

[andrea.passarella@iit.cnr.it](mailto:andrea.passarella@iit.cnr.it)

Scholar profile

<https://scholar.google.com/citations?user=sesKnygAAAAJ>



**Marco Conti**

[marco.conti@iit.cnr.it](mailto:marco.conti@iit.cnr.it)

Scholar profile

<https://scholar.google.com/citations?user=KniFTD0AAAAJ>



**Chiara Boldrini**

[chiara.boldrini@iit.cnr.it](mailto:chiara.boldrini@iit.cnr.it)

Scholar profile

<https://scholar.google.com/citations?user=rCrzxbsAAAAJ>