

**Alma Mater
Università di Bologna**
Dipartimento di
Matematica
Piazza di Porta S. Donato 5
40126 Bologna
Italy

Gabriele Sicuro
Associate Professor



ALMA MATER STUDIORUM | DEPARTMENT
UNIVERSITÀ DI BOLOGNA | OF MATHEMATICS

25 February, 2025

Large language models e dati sintetici per la teoria dei numeri
Progetto FAIR Spoke 8

English below.

L'intelligenza artificiale sta trasformando profondamente il nostro mondo, in un processo che ricorda la rivoluzione industriale. Al centro di questa trasformazione ci sono i modelli linguistici avanzati (Large Language Models, LLMs) che, grazie all'uso di database sintetici, stanno rivoluzionando il modo in cui affrontiamo problemi complessi e generiamo nuova conoscenza. Questa combinazione innovativa rappresenta un punto di svolta senza precedenti, capace di produrre risultati che non provengono dal cervello biologico umano o animale, ma da sistemi computazionali avanzati. Esempi di questo impatto straordinario includono l'elaborazione del linguaggio naturale e molte altre applicazioni.

Negli ultimi anni, l'IA basata su LLM e dati sintetici ha mostrato il suo straordinario potenziale anche nella scoperta matematica. In problemi complessi, come la classificazione delle curve algebriche, alcune congetture individuate grazie all'IA sono state successivamente dimostrate da matematici umani, sottolineando il ruolo cruciale di queste tecnologie nell'accelerare il progresso scientifico. Questo progetto ha come obiettivo il miglioramento delle nostre attuali capacità di fattorizzazione dei numeri interi nei loro fattori primi, una sfida che tocca sia la struttura teorica della teoria dei numeri che le sue numerose applicazioni pratiche. La sua rilevanza risiede nel valore della scoperta e della verifica di primi estremamente grandi, compiti che rimangono computazionalmente molto complessi.

La metodologia proposta combina tecniche avanzate di elaborazione del linguaggio naturale con macchine di Boltzmann, modellizzazione matematica e l'uso di dati sintetici per aumentare velocità e accuratezza nell'identificazione dei numeri primi. Questo progetto mira a sviluppare un quadro di ricerca robusto e scalabile, capace non solo di far progredire la teoria dei numeri primi, ma anche di dimostrare l'impatto trasformativo dell'intelligenza artificiale nella scoperta matematica. Le attività previste ai fini del progetto comprendono la generazione del database e l'addestramento del LLM.

Il candidato ideale per il progetto proposto possiede una formazione solida sia nelle scienze dure che nella programmazione e nella tecnologia dei modelli transformers e LLMs.

Artificial intelligence is profoundly transforming our world in a process reminiscent of the Industrial Revolution. At the heart of this transformation are advanced language models (Large Language Models, LLMs), which, through the use of synthetic databases, are revolutionizing the way we tackle complex problems and generate new knowledge. This innovative combination represents an unprecedented turning point, capable of producing results that do not originate from the human or animal brain but from advanced computational systems. Examples of this extraordinary impact include natural language processing and many other applications.

In recent years, AI based on LLMs and synthetic data has demonstrated its remarkable potential in mathematical discovery as well. In complex problems, such as the classification of algebraic curves, some conjectures identified through AI have later been proven by human mathematicians, highlighting the crucial role of these technologies in accelerating scientific progress. This project aims to enhance our

current capabilities in factoring integers into their prime components, a challenge that touches both the theoretical framework of number theory and its numerous practical applications. Its significance lies in the value of discovering and verifying extremely large prime numbers, tasks that remain computationally very complex.

The proposed methodology combines advanced natural language processing techniques with Boltzmann machines, mathematical modeling, and the use of synthetic data to increase speed and accuracy in prime number identification. This project aims to develop a robust and scalable research framework, capable not only of advancing prime number theory but also of demonstrating the transformative impact of artificial intelligence in mathematical discovery. The project's planned activities include database generation and LLM training.

The ideal candidate for the proposed project has a strong background in both hard sciences and programming, as well as expertise in transformer model technology and LLMs.



Gabriele Sicuro, PhD

Finanziato dall'Unione Europea - NextGenerationEU a valere sul Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Istruzione e ricerca – Componente 2 Dalla ricerca all'impresa - Investimento Investimento 1.3 , Avviso D.D. 341 del 15/03/2022, dal titolo: Future Artificial Intelligence Research, codice proposta PE0000013 - CUP J33C22002830006