

# Statement of Purpose for Applying to the PhD Program in Computer and Control Engineering

Luca Cotti, June 2024

The rapid advancement of Machine Learning is unlocking unprecedented opportunities to extract insights from vast amounts of data, which can be critical to tackle societal challenges and improve quality of life. My motivation for pursuing a PhD is to gain in-depth knowledge of Machine Learning and Data Science, and to contribute to the development of innovative, real-world technologies that have a positive impact on society. I also believe that the research skills I will acquire in a PhD in designing experiments, analyze data, and draw meaningful conclusions are crucial for a successful and high-reaching career in Data Science, which I intend to pursue as either a researcher or industry professional.

My motivations stem from my master's thesis, "Study and Development of a Digital Twin for the Simulation of User-Created Automations in a Green Smart Home", where I created a digital twin of a home to track the energy consumption of appliances through trigger-action routines. I extracted information about the appliances' operation modes from raw energy consumption data using state-of-the-art data mining techniques. I then developed a web application that tracks energy consumption, simulates various automation scenarios, and offers energy-saving suggestions. This work introduced me to impact that data-driven solutions can have on energy sustainability and quality of life. It also allowed me to discover an affinity for research, as it resulted in a conference paper for the International Conference on Advanced Visual Interfaces 2024 (AVI 2024) and a journal article published in Future Internet.

While moving from Brescia to Turin would be a significant challenge, I believe that the resources and opportunities I would gain at the Politecnico di Torino would be invaluable for my research and professional development. Particularly, the possibility of being part of the SmartData@PolitTO research center is very appealing to me, as it would allow me to collaborate with leading researchers in the field of Data Science and Machine Learning.

My future goals are backed by a solid foundation acquired through my studies in Computer Science and Engineering at the University of Brescia, where I attended courses in Machine Learning, Data Mining, and Deep Learning techniques. My passion for these fields led me to participate in several Kaggle competitions, applying my knowledge to practical problems. I also completed courses on Operational Research and Optimization Algorithms, culminating in a group project that involved tuning the Kernel Search heuristic algorithm to produce high-quality solutions for the Multiple Knapsack Problem. I also have experience in Cybersecurity, having completed a research project analyzing the implications of Hybrid Work on the security and privacy of information in organizations. During my studies I balanced the coursework with the unique cultural program of the Collegio Universitario Luigi Lucchini. The seminars, laboratories, travels abroad, and social events allowed me to engage with a community of brilliant minds, teaching me effective communication, teamwork, and time management—all skills that will be invaluable in my doctoral studies.

One area of research I am interested in exploring during my PhD is the application of Natural Language Processing techniques to Cybersecurity, to gather contextual evidence for supporting security analysts against cyber threats. My interest in this topic stems from the incredible potential of Large Language Models in understanding structured and unstructured data, which could be used to detect and mitigate cyber threats that are nowadays powered by Artificial Intelligence, persistent, and ubiquitous. Considering the state of the art, valid contributions could include adapting existing log-based models to detect new types of attacks through zero-shot classification, or expanding the types of logs that can be analyzed to include network traffic, system calls, and other sources of data.

Another intriguing topic is optimization of transportation systems. This topic interests me because of the potential to reduce traffic congestion, improve urban planning, and promote sustainable mobility, which could have incredible impact on energy sustainability, quality of life, and environmental conservation. Interesting contributions to this field could include the application of Data Science and Big Data techniques to data gathered from black boxes inside cars to predict traffic congestion and pollution, optimize routes, or study the efficiency of internal combustion, hybrid, and electric vehicles over different terrains and driving conditions.