AEROSPACE ENGINEERING / ARCHITECTURAL, URBAN AND INTERIOR DESIGN / ARCHITECTURE, BUILT ENVIRONMENT AND CONSTRUCTION ENGINEERING / BIOENGINEERING / **DATA ANALYTICS AND DECISION SCIENCES** / DESIGN / ELECTRICAL ENGINEERING / ENERGY AND NUCLEAR SCIENCE AND TECHNOLOGY / ENVIRONMENTAL AND INFRASTRUCTURE ENGINEERING / INDUSTRIAL CHEMISTRY AND CHEMICAL ENGINEERING / INFORMATION TECHNOLOGY / MANAGEMENT ENGINEERING / MATERIALS ENGINEERING / MATHEMATICAL MODELS AND METHODS IN ENGINEERING / MECHANICAL ENGINEERING / PHYSICS / PRESERVATION OF THE ARCHITECTURAL HERITAGE / STRUCTURAL, SEISMIC AND GEOTECHNICAL ENGINEERING / URBAN PLANNING, DESIGN AND POLICY

Chair:
Prof. Pier Luca Lanzi

DOCTORAL IN DATA ANALYTICS AND DECISION SCIENCES

The Ph.D. program in Data Analytics and Decision Sciences (DADS) aims at training highly qualified senior data analysts and data managers capable of carrying out research at universities, international institutions, tech and financial companies, regulatory authorities, and other public bodies. The program stems from the cooperation between three departments: Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Dipartimento di Ingegneria Gestionale (DIG), Dipartimento di Matematica (DMAT), and the Center for Health Data Science at Human Technopole. It allows the enrolled students to work in a highly interdisciplinary environment with strong connections to international research centers and private companies. The program provides successful candidates with the opportunity to acquire a high degree of professional expertise in specific scientific and technological fields.

The program lasts three years: upon its successful completion and final exam, candidates will be awarded the title of Ph.D. in Data Analytics and Decision Sciences. The first year is devoted to the courses that build the broad competence and the solid interdisciplinary set of skills required by data analytics. The following two years focus on the development of the Doctoral thesis. Students must spend at least one semester in a research institution abroad, taking advantage of the network of international collaborations of the three departments involved in the program.

The program aims at breeding the next generation of data scientists who will tackle the challenges and the opportunities created by the increasing availability of the massive amount of data. These data scientists will be able to capture the relevant aspects of phenomena at play, develop adequate models, supervise the development of analytic pipelines, critically analyze the results, and support the technological transfer.

Data Analytics and Decision Sciences graduates are equipped with unique skills and advanced knowledge that open up career opportunities at universities, international research centers and institutions, R&D departments, regulatory authorities, financial institutions, tech companies, and other public bodies.

FACULTY BOARD

- Prof. Lanzi Pierluca (Coordinator)
- Prof. Azzone Giovanni
- Prof. Caiani Enrico Gianluca
- Prof. Ceri Stefano
- Prof. Di Angelatonio Emanuele
- Prof. Flori Andrea
- Prof. Guerini Massimiliano
- Prof. Ieva Francesca
- Prof. Mangiaracina Riccardo
- Prof. Matteucci Matteo
- Prof. Punzo Fabio
- Prof. Roveri Manuel
- Prof. Secchi Piercesare
- Prof. Spagnolini Umberto
- Prof. Tanelli Mara
- Prof. Tubaro Stefano
- Prof. Vantini Simone

CHARACTERIZATION AND DETECTION OF DISINFORMATION SPREADING IN ONLINE SOCIAL NETWORKS

Francesco Pierri – Supervisors: Prof. Stefano Ceri, Prof. Fabio Pammolli

In the last decade, online Social Networking Sites (SNS) have become a pervasive presence in our everyday life. Billions of people nowadays use platforms such as Twitter, Facebook and Instagram to share text messages, photos and videos with their friends, but also to consume and disseminate news articles from news media outlets and blogs

Recently, these platforms have witnessed an explosive growth of malicious and deceptive information. The research community usually refers to it with a variety of terms, such as disinformation, misinformation and most often false (or "fake") news, hardly reaching agreement on a single definition.

Following a major shift in news consumption habits towards online content, and a drop in trust towards traditional news outlets, the problem of false information circulating online has become crucial, especially during periods such as political elections or epidemics, when the prevalence of unreliable online information has severe backlashes in the real world.

One example is the on-going COVID-19 pandemic, as the world experiences an "infodemic", i.e., an overabundance of information including false and misleading content, which undermines medical efforts and governmental efforts to fight the disease. For example, misinformation about masks contributed to low adoption rates and likely increased disease transmission. As false vaccine narratives spread, vaccine hesitancy will make it difficult to reach herd immunity and prevent future outbreaks.

Several reasons explain the rise of such malicious phenomenon. First, barriers to enter the online media industry have dropped considerably and (dis)information websites are nowadays created faster than ever, generating revenues through advertisement without the need to adhere to traditional journalistic standards (as there is no third-party verification or editorial judgment for online news).

Second, human factors such as confirmation biases, algorithmic biases and naive realism have exacerbated the so-called echo chamber effect, i.e., the formation of homogeneous communities where people share and discuss about their opinions in a strongly polarized way, insulated from different and contrary perspectives.

Third, direct intervention that could be put in place by platform government bodies for banning deceptive information is not encouraged, as it may raise ethical concerns about censorship.

Therefore, ever since 2016 U.S. Presidential elections and UK Brexit Referendum, the research community has witnessed an explosion of interest around the issue of disinformation, misinformation and other sorts of false information spreading in social media platforms.

The combat against online mis/disinformation is challenged by: the massive rates at which malicious items are produced, and the impossibility to verify them all; the adversarial setting in which they are created, as sources of misleading content usually attempt to mimic traditional news outlets; the lack of gold-standard datasets and the limitations imposed by social media platforms on the collection of relevant data.

Most methods for "fake news" detection are carried out by using features extracted from the news articles and their social context (notably textual features, users' profile, etc); existing techniques are built on this content-based evidence, using traditional machine learning or more elaborate deep neural networks, but they are often applied to small, ad-hoc datasets which do not generalize to the real world.

Important studies, featuring large-scale analyses, have produced deeper knowledge about the phenomenon, showing that: false news spread faster and more broadly than the truth on social media; social bots play an important role as "super-spreaders" in the core of diffusion networks; echo chambers are primary drivers for the diffusion of true and false content.

In this thesis, I leverage a hybrid computer science and network science approach to tackle the problem of disinformation, a term I will use as a shorthand to indicate all sorts of false, misleading and potentially harmful information spreading in online social networks.

Gathering data from multiple social platforms, we study the mechanisms and the actors involved in the spread of disinformation and other malicious content on social media during relevant events such as political elections and the on-going COVID-19 pandemic.

We carry out a systematic comparison of reliable information, published by mainstream and traditional news websites, versus unreliable information conveyed by websites which have been repetitively flagged for sharing disinformation, misinformation, hoaxes, fake news and hyper- partisan propaganda.

We provide evidence of superspreaders of disinformation, i.e., influential users who are responsible for most of the disinformation shared online, and we unveil links with far-right communities, which oftentimes exploit fabricated information to push their agenda. At the same time, we show that reliable information accounts for the majority of news stories circulating online, and that disinformation has a small yet non-negligible online prevalence which can still influence individuals' opinions and feelings.

Combining data from Twitter and Facebook, we investigate the interplay between vaccine-related disinformation shared on Twitter and the vaccine hesitancy and uptake rates measured among U.S. regions, following the roll-out of the COVID-19 vaccination program. Building a regression model which takes into account demographics, socio-economic and political factors, we find a significant association between online disinformation and the vaccine outcomes. Further preliminary analyses show similar results for the Italian context.

Finally, drawing on the results of these quantitative analyses, we deploy a methodology to accurately classify news articles based on the interactions between users that take on platforms like Twitter. Based on the assumption that users share differently disinformation news rather than mainstream articles, thus shaping different diffusion patterns, we train and test off-the-shelf machine learning classifiers which are able to classify the veracity of a news article, without the need of looking at its content.

All in all, our results pave the way to a better understanding of the issue of disinformation spreading in online social media, and highlight the need for intervention by platforms and governments to address this issue in a timely fashion.