EMbeDS

Economics and Management in the era of Data Science 2019

THE WORKSHOP

26, 27 NOVEMBER 2019

Sant'Anna School of Advanced Studies

BOOK OF ABSTRACTS

NOVEMBER 26, 2019

EMbeDS recruits, Session I

❖ Caterina Sganga, Dirpolis Institute.

Title: *Intellectual property and data ownership: balancing exclusivity and access.*

Abstract: New technologies have triggered an expansion in the scope, duration and objects of intellectual property rights. At the same time, the increased proprietary control over inventions, creations and data have negatively impacted on users' rights and freedoms, access to knowledge and innovation, competition and consumer welfare, and – last but not least – the balance between current and future authors and inventors. Examples are abundant. They range from the controversial aspects of the most recent EU copyright reform to the distortions generated by the overprotection of works online, the failure of the EU Database Directive, the instances of patent misuses and the disproportionate costs generated by patent litigation, the unharmonized and uncertain regulation of gene and biotech patents, the anticompetitive effects of data exclusivity in the field of pharmaceutical patents – and the list may continue.

One of the main threads of my research is the definition of the optimal balance between exclusivity and access in the design and implementation of intellectual property rights. I will take this opportunity to offer you an overview of some of my most recent papers on the matter, and a short presentation of the H2020 project "reCreating Europe", starting next January, in which we act as consortium coordinator and focus on the challenges of digital copyright and the digital single market as engines for creativity, cultural diversity and access to culture in Europe.

❖ Chiara Seghieri, Institute of Management.

Title: *Assessing access, quality and equity of care using observational data: current research and future directions.*

Abstract: Health policy and outcome research involves measuring and assessing access, costs, effectiveness, quality, and value of health care delivered in the everyday practice. In this talk, I introduce several health policy research questions and show different approaches to making inference on the basis of observational data. The goals of the research are to provide insights regarding how to reduce unwarranted variations in clinical practice, to organize, and deliver high quality care. An overview of future research developments that can be used in the quest for better population health will be also discussed.

- Andrea Vandin, Institute of Economics and Department of Applied Mathematics and Computer Science, DTU Technical University of Denmark.

Title: *Speeding up simulation of large-scale dynamical networks by aggregation.*

Abstract: Dynamical systems across many branches of science and engineering can be described in terms of massively many entities (possibly) interacting over large-scale networks. Examples are social networks involving spreading (or diffusion) of rumours, opinions, epidemics, malware, or technologies, as well as models of protein-interaction networks, and distributed computing systems.Mathematically, these models can often be described as stochastic processes or differential equations which, in all but special cases, do not enjoy analytical solutions. This requires computationally-expensive analysis based on stochastic simulators and numerical solvers, hindering our capability of dealing with complex large-scale real-world models. This talk will overview an array of recent techniques and tools of mine that tame the complexity of complex large-scale models by aggregating their constituting systems of equations while preserving the models' intelligibility and dynamics.

EMbeDS recruits, Session II

- Annamaria Murante, Institute of Management.

Title: *Co-producing services and health.*

Abstract: How making on air efficient, effective and structured co-production mechanisms in health sectors? In public administration co-production has been obtaining large attention at the policy making level (upperstream), designing and activating co-partecipation processes. However, still more work is to do at the operational level (midstream), in order to transform consumers in co-producers. To this end, the public administration science can learn a lot from service management and behavioral science. Starting from the Osborne's definitions of co-production(s), this work puts the attention on the co-design and co-construction processes. In details it shows two examples about how effective co-production mechanisms can be activated (i) when service quality improvement processes are also based on continuous and integrated data from co-producers' experience as well as (ii) when all stakeholders interested in the value to co-create are simultaneously and continuously involved in the service delivery and their involvement makes the outcomes co-produced lasting over time.

- Francesco Lamperti, Institute of Economics.

Title: *Climate risks in complex economies: are there opportunities beyond costs?*

Abstract: This talk provides a perspective on the effects that climate change might exert on our economies and, relevantly, on the opportunities that timely but cautiously climate policy could give. I build on both empirical and model-based evidence to argue that while

economic impacts risk to be extremely large and unevenly distributed across time, space and income classes, combating climate change is not just a matter of avoiding such costs. Specifically, technological change along certain trajectories has the potential to spur sustainable and inclusive growth. To reach the objective, policy should act across a variety of actors and domains. To conclude the talk, a research agenda emerging from the evidences previously discussed is proposed.

❖ Paola Cantarelli, Institute of Management.

Title: *Behavioral public management: A focus on work motivation in mission-driven jobs.*

Abstract: Research into the motivational bases of mission-driven professions is receiving unprecedented attention from both social scientists and practitioners. Thanks to its comprehensiveness, self-determination theory has rapidly become one of the most established motivation theories and is rapidly gaining hegemonic status across the social and behavioral sciences. At its core, it features the distiction between psychological needs in the workplace (i.e., autonomy, competence, and relatedness) and the distinction between extrinsic and intrinsic motivation. Further, self-determination theory allows reconciling and integrating the construct of public service motivation, which is native to public management, with analogous or complementary concepts - such as prosocial motivation - from related disciplines. We put self-determination theory to experimental tests with about 11.000 professionals working in government and healthcare organizations. Several discrete choice experiments illuminate the simoultaneous and independent effects of the satisfaction of psychological needs and the motivational forces on job choice.

❖ Daniele Giachini, Institute of Economics.

Title: *Market Learning.*

Abstract: This talk will provide insights on the relationship between markets and learning. The starting point is a simple model of financial market in which two agent repeatedly exchange two long-lived assets. The model is stochastic, in the sense that agents' wealth shares evolve depending on the realizations of a Bernoulli process. Agents do not know the true probability driving the process and assign to it heterogeneous and biased subjective probabilities. In such a context, the evolutionary wealth dynamics is equivalent to the weight updating of a fictitious sub-bayesian learner (that is, the market) who tries to understand the true probability combining agents' beliefs. Moreover, asset prices turn out to be the learner's estimates. The learning performance of the market can be considered superior: either it select the best subjective probability in the market, or it aggregates individual models in such a way that it is more accurate than any agent. In the end, a simplified version of the model can be used as an efficient model aggregation procedure in generic binary classification problems.

Scientific sessions organized by young EMbeDS researchers

1. New perspectives on climate change economics and big data

- Guido Cervone, Department of Geography & Institute for Cyber Science. The Pennsylvania State University.

Title: *Remote sensing target recognition using image spectroscopy and deep learning.*

Abstract: The current state-of-the science approach to hyperspectral imagery collection and exploitation derives from algorithms and tools developed in early 1990s, which apply many simplifying assumptions or expedient processing steps. The most significant simplification occurs in atmospheric correction where a single geometric solution for all elements of the radiance equation is applied to every pixel of a spectral image. We know this solution to be expedient, but also, error inducing. Another significant shortcoming is the inability of current algorithms to exploit more than one spectral image at a time. Atmospheric characterization and correction, and target detection is performed one image at a time: every image is an island.

These concepts and solutions have demonstrated operational success and their methods are ensconced in existing algorithms and software. While these existing solutions are based on sound physics, mathematics, and statistics, they lag behind the revolution in computational power, artificial intelligence, and agile sensors and platforms. Most importantly, they have been shown to fail under varied environmental conditions, obscuration, and target material conditions. It can be said that today's analysis is effective in performing material identification in optimal collection conditions. It can also be said that in non-optimal conditions, cloudy scenes, shadows, intimate mixtures of materials, liquid spills and residues, or any combination of the above, material identification of solids, liquids, or gases is ineffective, unrepeatable, or subject to unknown levels of uncertainty.

The gap in effectiveness stems from the incomplete solution of the radiance equation that a single hyperspectral image and existing analysis methods afford. This talk discusses new research that promises to greatly improve today's state-of-the-art hyperspectral imagery analysis methods by expanding the spatial and temporal dimensions of the radiance equation solution and implementing Deep Learning.

- Vantini Simone, MOX - Department of Mathematics. Politecnico di Milano.

Title: *Object-Oriented Data Analysis for Climate Change and Energy Economics*.

Abstract: The vast and systematically increasing availability of storage and computational power we are experiencing in the latest years is bringing a real revolution in terms of analytical, inferential and forecasting tools that are now available to scientists in every field of knowledge. climate change and energy economists are no exception. New statistical learning methods, enabled by such computational power, allow for novel questions to be

asked to novel data sources, and thus to obtain new and more insightful conclusions. On this line of thought, we will give a brief introduction to object-oriented data analysis (OODA), a new statistical framework explicitly aimed at modelling complex data sources. In this modeling context, the classical concept for scalar or multivariate random variable is indeed replaced by the concept of object random variable. Consequently, in OODA the typical data set is not made of numbers or Euclidean vectors, but a collection of complex mathematical objects (e.g., functions, images, networks) embedded in a suitable mathematical space meant to formalize application-specific relations between sample units. Recent applications of OODA techniques in different and many fields of science are countless. Nevertheless, very few applications can be found in the field of Climate Change and Energy Economics, thus pointing out the existence of an unexploited potential of this type of techniques in these two fields. With respect to this discrepancy, after a gentle introduction to OODA, this talk will showcase some recent applications in which state-of-the-art OODA techniques have been fruitfully used at MOX, the Modeling and Scientific Computing Lab of the Department of Mathematics of Politecnico di Milano. In detail, we will focus on the assessment of the impact of a behavioral intervention on energy consumption performed by a major energy utility in a city in the southern Italy; the forecasting of daily demand and supply curves for natural gas trading; and global, time-variant sensitivity analysis for an integrated assessment model ensemble.

❖ Chiara Io Prete, Department of Energy and Mineral Engineering. The Pennsylvania State University.

Title: *California's cap-and-trade program and emission leakage in the electricity sector: an empirical analysis.*

Abstract: We conduct the first econometric analysis of leakage in the electricity sector from California's cap-and-trade program. The paper presents three sets of empirical results that support the hypothesis of leakage. First, we measure the policy impact on baseload power plant operations in the Western Inter-connection applying a differences-in-differences estimator to a novel dataset at the monthly level from 2009 to 2016. Second, we preprocess the data to improve balance between treated and control plants by matching on hour-of-day specific variables, and explore treatment effect heterogeneity across day-time and nighttime hours using daily measures of plant utilization.

Third, we test for leakage from the cap-and-trade program by examining the relationship between emission allowance prices and scheduled power imports into California.

Results suggest a policy-induced reduction in natural gas combined cycle generation in California and an increase in coal-fired generation in the Western U.S., corresponding to a leakage rate of about 70%.

2. The value of big data in healthcare

❖ Bin Yu, Departments of Statistics & Computer Science. University of California Berkeley.

Title: *Iterative Random Forests (iRF) with applications to biomedical problems*.

Abstract: Genomics has revolutionized biology, enabling the interrogation of whole transcriptomes, genome-wide binding sites for proteins, and many other molecular processes. However, individual genomic assays measure elements that interact in vivo as components of larger molecular machines. Understanding how these high-order interactions drive gene expression presents a substantial statistical challenge. Building on random forests (RFs) and random intersection trees (RITs) and through extensive, biologically inspired simulations, we developed the iterative random forest algorithm (iRF) to seek predictable and stable high-order Boolean interactions. We demonstrate the utility of iRF for high-order Boolean interaction discovery in two prediction problems: enhancer activity in the early Drosophila embryo and red hair phenotype using UK BioBank data. The latter is a proof-of-concept step towards suggesting gene variants behind cardiovasuclar phenotypes for single cell experiments as part of a Chan-Zuckerberg Biohub Intercampus Award to UC Berkeley, UCSF and Stanford. Finally, a connection is made between iRF and our PCS framework for veridical data science (PCS stands for predictability, computability and stability).

❖ Francesca Ieva, MOX - Department of Mathematics. Politecnico di Milano.

Title: *The broken promise of Big Data in Healthcare and the new challenges of Health Analytics.*

Abstract: The term “Big Data" in healthcare refer to a huge and widely heterogeneous amount of data arising from current practice in any clinical area. In the last decade, "Big data" was supposed to change the way clinicians, researchers and people in charge with healthcare government carry out their work and think about their mission. Was this revolution real as a matter of facts? Despite the healthcare community is starting using data and real world evidence in general to support decisions, to develop guidelines and tools for clinical practice and practitioners, to monitor postmarket safety and adverse events, and to make regulatory decisions, many practical and ideological barriers still remain. In this talk, we will discuss some broken promises of the Big Data in healthcare, in order to point out the new challenges that a smart health analytics should pursue for addressing personalized medicine issues. In such a context, a key role is played by the new skills required to the data scientists for handling complexity of the data.

3. Prediction, learning and inference: data science from theory to practice

❖ Nicolò Cesa-Bianchi, Dipartimento di Informatica e Data Science Research Center, Università degli Studi di Milano.

Title: *Nonstochastic Bandit Problems on Graphs.*

Abstract: Multiarmed bandits are a class of sequential decision problems that have been successfully used to model applications ranging from routing in networks to online advertising. Bandit algorithms, characterized by a trade-off between exploitation and exploration, are often analyzed in terms of sequential learning with partial feedback, a broader setting where the learner has a costrained access to the environment. These information constraints can be naturally expressed using graphs. In this talk we will revisit nonstochastic bandits problems using two conceptually different graph-based information models. In the first one, actions are nodes in a graph, and the learner's feedback in each decision round is the payoff of all actions in the out-neighborhood of the action chosen at that round. In the second model the learning agents are nodes in a communication network, which they use to cooperatively solve a bandit problem. In both cases, we study the extent to which observation and communication, measured through graph-theoretic quantities, improve the minimax learning rate of the underlying bandit problem.

❖ Michel Lechner, Swiss Institute of Empirical Economic Research, University of St. Gallen.

Title: *Modified Causal Forests for Estimating Heterogeneous Causal Effects.*

Abstract: Uncovering the heterogeneity of causal effects of policies and business decisions at various levels of granularity provides substantial value to decision makers. This paper develops new estimation and inference procedures for multiple treatment models in a selection-on-observables framework by modifying the Causal Forest approach suggested by Wager and Athey (2018) in several dimensions. The new estimators have desirable theoretical, computational and practical properties for various ag-gregation levels of the causal effects. While an Empirical Monte Carlo study suggests that they outper-form previously suggested estimators, an application to the evaluation of an active labour market pro-gramme shows the value of the new methods for applied research.

❖ Christian Brownlees, Department of Economics and Business, Universitat Pompeu Fabra.

Title: *Community Detection in Partial Correlation Network Models.*

Abstract: We introduce a class of partial correlation network models with a community structure for large panels of time series. In the model, series are partitioned into latent groups such that correlation is higher within groups than between them. We then propose

an algorithm that allows us to detect the communities using the eigenvectors of the sample covariance matrix. We study the properties of the procedure and establish its consistency. The methodology is used to study real activity clustering in the U.S..

NOVEMBER 27, 2019

- Bin Yu, Departments of Statistics & Computer Science. University of California Berkeley.

Title: *Veridical Data Science.*

Abstract: Building and expanding on principles of statistics, machine learning, and the sciences, we propose the predictability, computability, and stability (PCS) framework for veridical data science. Our framework is comprised of both a workflow and documentation and aims to provide responsible, reliable, reproducible, and transparent results across the entire data science life cycle. The PCS workflow uses predictability as a reality check and considers the importance of computation in data collection/storage and algorithm design. It augments predictability and computability with an overarching stability principle for the data science life cycle. Stability expands on statistical uncertainty considerations to assess how human judgment calls impact data results through data and model/algorithm perturbations. We develop inference procedures that build on PCS, namely PCS perturbation intervals and PCS hypothesis testing, to investigate the stability of data results relative to problem formulation, data cleaning, modeling decisions, and interpretations.
Moreover, we propose PCS documentation based on R Markdown or Jupyter Notebook, with publicly available, reproducible codes and narratives to back up human choices made throughout an analysis.
The PCS framework will be illustrated through two neuroscience projects:
- movie reconstruction from brain signals, and
- DeepTune for charaterizing neurons in primary visual cortex V4.

PERSPECTIVES FROM INSTITUTIONAL STAKEHOLDERS

- Giovanni Felici, European Research Council Executive Agency and Consiglio Nazionale delle Ricerche.

Title: *The European Research Council Grants: funding opportunities for frontier science*.

Abstract: After its first ten years of activity, ERC has confirmed its worldwide recognition as one of the most effective funding agencies for bottom-up, creative high-risk / hi-gain research in all fields of science. ERC current set-up remains unchanged in the forthcoming Framework Program, thus confirming the ERC prominent role for European Research. In this talk we will recall the basic principles on which ERC grants are based; how Starting, Consolidator, Advanced and the more recent Synergy ERC grant schemes work, what are the conditions to apply, and discuss the main ingredients of a successful proposal.

Additionally, we will focus on the landscape of funded projects related to Data Science, providing an overview on the ERC-funded research in this field. Finally, time will be devoted to discussion and to address questions from the audience.

- Fosca Giannotti, Istituto di Scienza e Tecnologie dell'Informazione, Italian National Research Council (CNR), Pisa.

Title: *Explainable Machine Learning for Trustworthy Artificial Intelligence.*

Abstract: Black box AI systems for automated decision making, often based on machine learning

over (big) data, map a user's features into a class or a score without exposing the reasons why. This is problematic not only for the lack of transparency, but also for possible biases inherited by the algorithms from human prejudices and collection artefacts hidden in the training data, which may lead to unfair or wrong decisions. The future of AI lies in enabling people to collaborate with machines to solve complex problems. Like any efficient collaboration, this requires good communication, trust, clarity and understanding. Explainable AI addresses such challenges and for years different AI communities have studied such topic, leading to different definitions, evaluation protocols, motivations, and results. This lecture provides a reasoned introduction to the work of Explainable AI (XAI) to date, and a quick surveys the literature on machine learning and symbolic AI related approaches. I focus on the urgent open challenge of how to construct meaningful explanations of opaque AI/ML black-box decision systems, introducing our early results on the local-to-global framework as a way towards explainable AI.

- Denise Amram, Sant'Anna School of Advanced Studies.

Title: *Predictive Jurisprudence: an Interdisciplinary Research*.

Abstract: Research is everyday more data-intensive. From this perspective, legal studies based on case-law analysis become a suitable domain to apply Machine Learning techniques in order to contribute to the scientific legal debate from a completely innovative perspective. In this context, Predictive Jurisprudence aims at developing and training an algorithm within the lower Italian case-law on three selected case-studies. The first one refers to the alimony in case of divorce: queries are pre-determined by law, but their judicial interpretations continuously evolve. At this regard, current reform bills are proposing to introduce new criteria, whose efficacy could be discussed in light of our analysis. The second pilot and the third one develop within the Observatory on Personal Injury Damage studies. The algorithm may contribute to the identification of criteria for non-pecuniary losses compensation beyond the current interpretations and attempts to standardize a head of damage which is strongly related to the individual subjectivity, like the so-called pecunia doloris. Within this core-analysis, the algorithm could be better-

trained to explain non-pecuniary losses in case of burn out, whose boundaries are still discussed both from clinical and legal perspectives. The involvement of interdisciplinary expertise (including text mining, social data-mining, medico-legal, psychiatrics, etc.) will firstly impact on a better understanding of the judicial reasoning and, more generally, of a given legal system, but it also opens to undiscovered developments in terms of policy-making.

❖ Paolo De Rosa, Italian Digital Team.

Title: *How the digital transformation in the public sector may help to boost growth through digital infrastructures, shared platforms and human-centred services.*

Abstract: In the last year, significant changes in digital services, data and infrastructures of the public sector have been introduced that might contribute to reshape the role of government in the growth of the country. A description of the main challenges encountered, what has been successful, what still needs to be done and major lessons learned in the course of the digital transformation journey in the public sector.

PERSPECTIVES FROM PRIVATE STAKEHOLDERS

❖ **Piero Altoè**, Business Development Manager South-Europe and Middle-East, NVIDIA.

Title: *When accuracy is critical: the case of deep learning at scale*.

Abstract: The compute needs of AI researchers continue to increase as the complexity of DL networks and training data grow exponentially. Training in the past has been limited to one or a few GPUs, often in workstations. Training today commonly utilizes dozens, hundreds or even thousands of GPUs for evaluating and optimizing different model configurations and parameters. Also, the most complex models require multiple GPUs to train faster or support larger configurations. In addition, organizations with multiple AI researchers need to train many models simultaneously, requiring extensive compute resources. Systems at this massive scale may be new to AI researchers, but these installations have traditionally been the hallmark of the world's most important research facilities and academia, fueling innovation that propels scientific endeavor of almost every kind. The supercomputing world is evolving to fuel the next industrial revolution, which is driven by a re-thinking in how massive computing resources can come together to solve mission critical business problems. NVIDIA is ushering in a new era where enterprises can deploy world-record setting supercomputers using standardized components in months or even weeks. Designing and building computers at these scales requires an understanding of the computing goals of AI researchers in order to build fast, capable, and cost-efficient systems. Developing infrastructure requirements can often be difficult because the needs of research are often an ever-moving target and AI models, due to their proprietary nature,

often cannot be shared with vendors. Additionally, crafting robust benchmarks which represent the overall needs of an organization is a time-consuming process. It takes more than just a large GPU cluster to achieve the best performance across a variety of model types. To build a flexible system capable of running a multitude of DL applications at scale, organizations need a well-balanced system which at a minimum incorporates:A low-latency, high-bandwidth, network interconnect designed with the capacity and topology to minimize bottlenecks. A storage hierarchy that can provide maximum performance for the various dataset structure needs. These requirements, weighed with cost considerations to maximize overall value, can be met with by the design presented in this talk.
Ref.
https://devblogs.nvidia.com/dgx-superpod-world-record-supercomputing-enterprise/
https://blogs.nvidia.com/blog/2019/06/17/dgx-superpod-top500-autonomous-vehicles/

❖ Paolo Carnevale, Executive Director, Fondazione ENI Enrico Mattei.

Title: *Climate change, decarbonization of energy systems and WEF nexus: the role of Big Data.*

Abstract: The effects of climate change currently underway on terrestrial ecosystems represent a momentous challenge for our generation and for the future ones, requiring the individuation of effective and sustainable solutions in the short term. Required solutions must be adaptable to different latitudes and result from a renovated paradigm of global governance. If, on one side, scientific debate on the issue has been flourishing for at least three decades, on the other, policymakers and public opinion's perception of the urgency and complexity of climate change emerged very lately. In this framework, technology plays a crucial role in the acceleration of processes for decarbonization of energy-intensive production systems, particularly within sectors mostly responsible for CO2 and greenhous gases' emissions: energy production and distribution, heavy industry, transport, buildings and land use activities. The latter, and its related supply chains, are responsible for 24% of yearly emissions at global level. Hence, along with traditional analysis of energy-intensive sectors, further investigations on the impact of land use activities – and in particular of the Water-Energy-Food nexus – apppear essential to pursue a global framework characterized by economic and social development in coherence with environmental protection. Today, new research approaches based on the utilization of big data and citizen science are able to provide data and quantative indicators and subsequently great support to decision- and policy-making for the responsible management of natural resources, as well as for the individuation of policies for sustainable economic development.
Two explicative cases of data science application are hereinafter presented. Both cases refer to an increasingly pivotal region for the energy balance and its relative impact on climate change at global level, such as Sub-Saharan Africa.

a) The creation of a harmonized high-resolution dataset for electricity access and use tiers within an open-source geospatial computing framework for the development of a geospatial model for electrification focused on the promotion of productive uses of energy; b) The development of an operative webGIS dashboard for resource-management able to harmonize hydrogeological, climate-related and evironmental risks, with particular reference to WEF nexus.

❖ Antonio Pelliccia, Business Development Healthcare & Life Science, IBM Italia.

Title: *AI in Healthcare & Life Science: examples and challenges.*

Abstract: "The explosion of clinical and research data, together with the increased computational capacity provided by high-performance-machine, have set a favourable environment for the renaissance of Artificial Intelligence or, more in general, of new technologies to improve efficiency and support the healthcare and life science industries. AI is leading the way with Machine Learning and Deep Learning to analyze vast amount of data and obtain new insights from drug development, to real world evidence, from clinical research to health outcome and disease prevention. AI (and not only) is at the top of the hype, but there are still important challenges to be addressed and the 'general AI' is far from the reality. At the contrary, when tasks are simple and repetitive, AI and other technologies can speed up the job, assist the humans with additional insights, reduce clinical risk or suggest the research pathway for new drug development. Examples are already present in healthcare and life science: chatbot can interpret medical language and reply with proper scientifical documentation, genes or mutations can be analysed in minutes and suggest proper and personalized oncology therapies, software tools can analyse real time radiological images and detect intracranial hemorrhage, algorithms can predicts an IC50 sensitivity value by analysing molecular structure of compounds, transcriptomic profiles of cancer cells as well as prior knowledge about interactions among proteins within cells. As AI advances, and humans and AI systems increasingly work together, it is essential that we trust the output of these systems to inform our decisions. Alongside policy considerations and business efforts, science has a central role to play: developing and applying tools to wire AI systems for trust, fairness, robustness, transparency and explainability."

❖ Gian Luca Cattani, R&D Director and Member of Board of Directors, MAPS Group.

Title: *Data Driven Governance: Combining Administrative and Clinical Information for a Better Healthcare*

Abstract: "Quality healthcare is the result of the combination of many different factors. Clinical professionals excellence as well as state-of-the-art infrastructures are not sufficient elements for high quality healthcare, even less so if this needs also to be economically sustainable. The entire process chain of healthcare from the origination of the demand, to

the delivery of services must be properly considered: an intelligent and careful governance of healthcare processes improves the quality of care, enhances the trust of patients in healthcare institutions, makes an efficient usage of resources, and reduces costs. In this seminar, we will highlight the importance of leveraging clinical information to best inform governance and administrative decisions in healthcare. In particular, we will see how this entails the need for algorithmic tools to automatically retrieve such information from clinical data that are often stored as narrative text, e.g., in referral requests and reports, or in discharge letters. This information combines with administrative structured data to provide a complete picture of healthcare demand and of its impact on healthcare service providers. If properly actioned, such knowledge improves pervasively the decision making process of healthcare services managers, and of healthcare professionals. Concrete examples taken from our experience as software solution providers for healthcare authorities and hospitals will help us proving our point."