ISSN 2724-6132 / 2022-1

Science & Technology

FORESIGHT

from society to research

Report of the workshop:

"The Complexity of Artificial Learning"

held online on 05/05/2021 and 21/10/2021

WG COMPLEXITY

WG MADFUTURE

National Research Council of Italy

Report of the workshop
“The Complexity of Artificial Learning”

held online on 05/05/2021 and 21/10/2021

ISSN2724-6132 /2022-1

Table of Contents

INTRODUCTION

Artificial learning will significantly affect human life and science in the next decades. However, its basic principles of functioning are still theoretically not well understood. The aim of the workshop was to examine the relations and perspectives of the interaction between Data and Computer Sciences and Complexity theory on this matter.

Six world-renewed experts have been invited to debate the topic of artificial learning, enucleating a set of basic issues where cross-disciplinary interactions among Data, Artificial Intelligence, and Complexity are expected in the coming years. The following questions have been posed to the experts to ignite the debate:

1. What are the current limitations of inference from data and learning? Is it possible to go beyond them?
2. How to use machines to find emergent organisations in our vast data sets? Can they be integrated into the scientific process? Can they be used to automate simulations of physical, chemical, or biological processes?
3. How to account for variations due to new or unexpected conditions in an open and changing environment, like for instance data non-stationarity?
4. Can we gather insights into the plasticity and the learning process/phenomenon of deep learning models?

EXECUTIVE SUMMARY

The following conclusions could be initially drawn from the discussion so far, thus raising additional questions.

Learning systems work in an *unreasonably effective*[1] *[1]* manner when confronted with well-defined, well-posed problems, with “simple” metrics of performance.

This poses a series of questions on their functioning. For instance, why do learning systems not suffer from over-parameterization even though the canonical theory of inference would claim so? There is evidence that learning machines should be seen as out-of-equilibrium systems. What does it mean from the point of view of training protocols? What are the implications of such a statement? Can this perspective open to a new understanding of the learning process?

We may argue how those systems could be included in the scientific process. This would entail a reconsideration of the scientific work, by upgrading the scientific method dating back to Galileo. New approaches based on the use of data should be suitably (and critically) considered and tailored. What are the intrinsic limitations?

[1] Such expression is used by the author in [1], and herein used to ignite the discussion.

Who should read this report

This preliminary report is directed to scientists, policymakers, and more generally to those working in the field of Science and Technology Policy. Nevertheless, we would like to point out that the main focus of the workshop that led to this report is on the scientific aspect of the topic 'Complexity of Artificial Learning', with minor emphasis on the technological and policy dimensions, as put forward by the Science and Technological Foresight Project.

Nonetheless, we would like to shortly report herein both concerns and current strategies in the field of AI, starting from the two main points coming out of the '*European approach to trustworthy AI*'. Two main issues are at the core of such an approach the **responsibility** of developers and AI-powered systems and the need for increased **accuracy** of such systems. The former should be read in wide terms, meaning that the sustainable and ecological responsibility of AI systems must be encouraged[2], but also that a clear and agreed mechanisms are needed to ensure liability and accountability for AI systems, their outcomes and their developers. Legally, responsibility cannot be replaced by any certifications. Regarding accuracy, it must be intended as the ability of the system to form a correct judgement, mitigating unintended risks. Explainability, discussed later in this report, must be considered in this context because it can reduce accuracy, or vice versa (larger accuracy may bring lower explainability).

We refer the interested reader to the EU reports in this regard, in particular to the '*Europe fit for the Digital Age'* programme[3], to the legal framework on AI[4], and to the '*Guidelines for Trustworthy AI*'[5]. Far from being exhaustive in putting here some pointers to other reports focusing on the aforementioned aspects, we refer the interested reader to the remarkable reports prepared by AlgorithmWatch.

[2] Independent High-Level Expert Group on Artificial Intelligence, European Commission. "Ethic Guidelines for Trustworthy AI", April 2019. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[3] Europe fit for the Digital Age, https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

[4] Proposal for a Regulation laying down harmonised rules on artificial intelligence , https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[5] Communication: Building Trust in Human Centric Artificial Intelligence, https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence

What are the current limitations of inference from data and learning? Is it possible to go beyond them?

Limitations of inference from data and learning

Learning processes need a very large amount of data on which preprocessing activities have typically been performed as e.g., labeling purposes. Such a task is crucial for any learning algorithms to be applied, raising the subject of (structured) data as a fundamental brick. Data preprocessing is a first - and sometimes rather severe - limitation of the approaches in use nowadays, which may prove impractical in some cases; to some extent, this confirms that (deep) neural networks (DNNs) are still far from mimicking the brain. DNNs have been successfully used in the context of very specific problems - well-posed, well-defined, with simple measures of performance – proving to be *unreasonably effective* [1]. DNNs can be described and characterized in terms of neurons, operations, etc., yet we understand only partially the inner learning mechanism and how new information emerges from the network. *A mathematical theory of deep learning would illuminate how high-dimensional spaces work, allowing us to assess the strength and weaknesses of different network architectures* [1]. Actually, there is no guarantee that such an approach may work *a priori* in every circumstance, which is a limitation of the current scope of their use [3].

Despite their growing successes, one should be aware that, to date, learning systems have shown no 'general intelligence', reasoning, ability to represent the world, consciousness, attention, or causality. This may impact the output, as well as biased data may negatively impact it.

Fundamental physical limits to inference from data

Forecasting, one of the most potentially interesting applications of artificial learning, is proving a very hard task as well [4]. In fact, the temptation of assuming that brute-force data analytics can substitute theory-building must be resisted. Indeed, modelling requires the ability to select the most relevant variables, a *skill* that DNNs still lack. In this data deluge, the idea that Big Data – so-called because of their velocity, variety, volume, value, and veracity – can be 'enough' to model a phenomenon seems not to be confirmed by reality at date. Take the case of deterministic systems: based on the premise that 'from the same antecedents follow the same consequences', an analogue must be found in the past to understand the future. But limitations to that are already known, as stated by physicists like Maxwell, Richardson, Lorentz. Indeed, finding an analogue in the case of a macroscopic system can require an exponentially large time [2].

Another factor to be considered in dynamic problems is the presence of multiple significant scales and a variety of degrees of freedom. General features and main 'ingredients' can be described by equations otherwise hidden when initially stating the first principles, and the former ones typically allow emergent features to appear and be recognised.

Progress and needed pieces to overcome the current limitations

Although several limitations exist - as those cited above, *unreasonable effectiveness* is there. In the next few years, great progress is expected in the following areas. In the field of reinforcement learning – a powerful approach based on a renewal system– strong advances have been demonstrated both in a lab and in real-world settings. Also, unsupervised learning, a class of algorithms in which data used for learning are not tagged nor labelled by humans before their use, is showing great promises as a step towards bypassing the costly generation of training data. In this context, self-supervised learning has produced astonishing tools such as Google BERT or OpenAI GPT-3 in the field of Natural Language Processing (NLP). On this aspect, the workshops have emphasised the potential of novel research lines that promise to limit (or even remove) the impact of preprocessing activities, such as self-supervised learning. Furthermore, other approaches, following different approaches from those used in BERT or GPT-3, i.e., those falling into the so-called knowledge computing, must be cited. The core idea is to somewhat replace huge neural networks with smaller ones fed by knowledge graphs, i.e., a way to represent knowledge by means of a network stored in a graph database. Knowledge computing seems to have the potential for supporting reasoning-based approaches, although further studies are required in this sense.

Transfer learning, i.e., the (re)use of a model trained on task A as a base to deal with task B, is an emerging field that has the potential of showing great results from learning on different domains and then reusing such knowledge in the domains of interest. Another fascinating approach is continuous learning, in which the model is continuously updated (retrained) thanks to the incoming stream of data (it is worth highlighting that, in non-continuous algorithms, the training phase is separated from the inference phase). On the topic of causality, the so-called invariance learning must be mentioned, for instance, used in computer vision to recognise e.g., an object even when its appearance changes because of e.g., translation, rotation, size, and so on. From a conceptual point of view, the ability to recognise something independently from something else (e.g., the background) has great potential and can provide huge benefits in this field. In some fields, the potential of data augmentation must be acknowledged, as later discussed in the case of clinical questions.

When considering longer time horizons, the need for new theories arises as a need for overcoming the current limitations of those systems. Again, it is *unreasonably effective* how DNNs, with lots of layers and weights, can generalize well also in the case of huge networks with a huge number of parameters, avoiding data overfitting as the canonical theory of inference would instead claim.

Towards a theory of learning?

Fundamental problems in understanding the functioning of neural networks (NNs) are known and unsolved since the end of the last century:

- The overfitting problem: back-propagation algorithms that are employed to train the machines, typically require fixing *millions* of parameters.
- Is there an effective number of such parameters?
- The main tool for training algorithms is (surprisingly!) simple, basically gradient descent (hard to believe for glass physicists). Why do such dynamics not head to local minima?
- How many training samples are needed for a given task? Usually, the learning procedure is on a given sample set - like e.g., CIFAR-10 with 50000 samples. Is there a minimum value for such samples?[6]
- If this is the case, are we close to such a minimum? If not, is it because of architectures or algorithms?

Complexity scientists are currently trying to get insights into these issues with the aim of constructing the theoretical foundations of learning. The starting point is the analysis of simplified models.

Is modelling bringing us to understand deep learning?

Minimal models (akin to the Ising model in statistical physics) have been of great help in understanding the learning process [6]. A simple example is the Teacher-Student perceptron [7] where the main parameter is the ratio $\alpha = n/d$ of the number of samples and the dimension of the network. It is possible to compute the lowest error that can be achieved with a given number of samples. The model also demonstrates a phase transition in the generalization error's dependence on the sample complexity.
A big mystery is that we usually deal with an NP-complete problem that would require too much computational effort to be solved in practice. Actually, the model also shows "hard" regimes of α where the problem cannot be solved in polynomial time. Another example demonstrating hard regimes is the phase retrieval problem, an inverse procedure in image processing where one would like to retrieve the phases from the measured intensities.

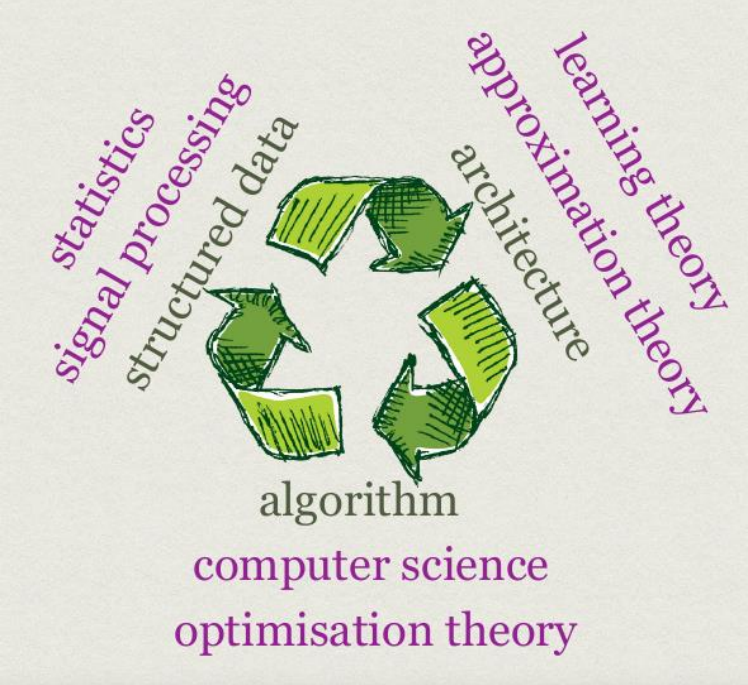
What steps towards a theory of deep learning?

Based on this body of research, it turns out that the issue is to understand the interplay of three main ingredients:

- Structured data
- Architecture
- Algorithms

They correspond to established fields of mathematics, computer science and statistics (see the scheme below), which however remained so far rather unrelated. Thus, it is a future challenge to integrate them into a common ground.

[6] There exists an approach to this problem called Computational Learning theory by Vapnik et al. [5]



In this respect, the importance of simplified models is demonstrated by the recent achievements on the above three topics:

- Structured data: from independent and identically distributed random variables to more realistic structured dataset
- Architecture: new tools to account for the many hidden layers that are present in DL networks
- Algorithms: message passing (MP) versus gradient descent (GD).

To give an idea of the theoretical achievements [4], it was possible to prove rigorously that over-parameterization helps reduce the parameter α for gradient-descent algorithms. Altogether, this demonstrates that over-parameterized neural networks need fewer samples to learn.

Learning is a non-equilibrium process

The need for novel theories is also demonstrated by the fact that the canonical theory of inference seems not to apply to Deep Learning models. For these models, the optimization problem is very high dimensional, with hundreds of parameters, and the loss functions are non-convex. Despite this, using learning algorithms suitable for convex problems (gradient descent) with various heuristics on top (SGD, ADAM/AdaGrad dropout, batch normalization, Glorot initialization, regularizations, …) ensures unreasonably good results. It appears relatively easy to avoid overfitting while still learning the whole training set.

To understand why this is possible, Google ran a totally empirical experiment in 2019, by training 10000 models over two image classification datasets (CIFAR-10, Street view numbers), under all combinations of hyperparameters and optimization [8]. They obtained a

large pool of models and evaluated each of them with 40 complexity measures. Among these metrics that could be related to the generalization ability of the network, the more stable one was a measure of the flatness of the error surface. This is an important observation as if you have a look at deep nets that are not convex, one will rather expect a very complex landscape, with a huge number of local minima, flat regions, and traps. However, the minimizers based on existing algorithms do not sample according to a Gibbs distribution. In other words, each model has a complex loss function, but all existing algorithms are not good samplers of the 0-1 (error) loss.

Altogether, these machines should be seen as out-of-equilibrium systems. In this direction, there is a lot to do, as the connection between learning and algorithmic behavior is not understood. It can be envisaged that this aspect will strongly benefit from the interaction with Complexity Sciences which usually deal with non-equilibrium dynamics.

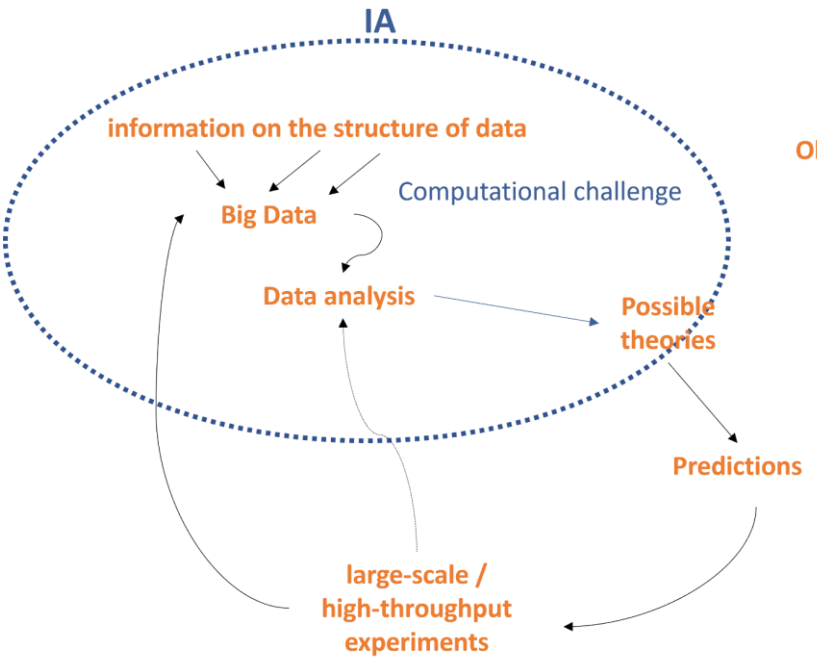
How to use machines to find emergent organizations in our vast data sets? Can they be integrated into the scientific process? Can they be used to automate simulations of physical, chemical, or biological processes?

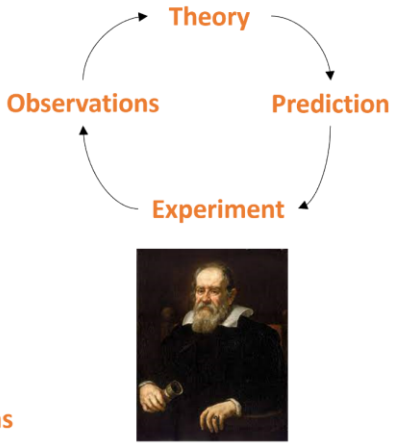
Today, the applications of Deep Neural models to science problems are very impressive and growing in several diverse fields:

- AI itself, as these machines are complex systems themselves
- DNN and PDEs
- Quantum physics
- Astrophysics
- Social sciences
- Information Retrieval

Some of the most interesting ones include AlphaFold, a revolutionizing model for the prediction of protein structure. Currently, AlphaFold 2 is approximating the structures with the accuracy of X-ray spectroscopy. There is a lot of work going on also on Deep Neural Networks to solve Partial Differential Equations. In the beginning, the approach was based on supervised learning based on time series modelling. Recently, the strategy is to learn operators by sampling in an unsupervised approach. There are promising results that are expected to be useful for instance for material science.

There are thus good reasons to believe that in the next future we might include these methods in the scientific process. We should integrate the approach, which dates to Galileo's times, by adding another component of the scientific method, relying on the use of data (see the scheme below).





Can we obtain models from data?

Despite the many positive results, the integration of DNN learning in the scientific process cannot be considered a straightforward procedure. Indeed, brute-force correlation analysis cannot be compared with a summary of scientific knowledge in a few equations, which are compositions of elementary laws. The combinatorial and compact nature of the equations is much richer than brute-force data correlation analysis.

Although one can think of machine learning as an innovative tool to solve given problems. But one may face an even more ambitious task, namely whether it is possible to use only data as inputs for model building. On general grounds, any procedure aiming at extracting effective models from data should face two main difficulties:

- Typically, we do not know *a priori* the proper variables that should enter the effective model and not even how many they are
- Even in the (lucky) case we know the proper variables, if the dimension of the “phase space” of the system under study is larger than 5 or 6, it is impossible to find analogs, and the protocol collapses.

Also, there is a subtle relation between data and models. The concept of algorithmic complexity contributed to making mathematically rigorous the long-standing idea that, to understand empirical phenomena, it is necessary that the rules which generate the data are “simpler” than the data itself. A key issue for the appraisal of the relation between algorithmic complexity and algorithmic learning has to do with a much-needed clarification on the related but distinct concepts of compressibility, determinism, and predictability. For instance, the evolution law of a chaotic system is compressible, but a generic initial condition for it is not, making the time series generated by chaotic systems incompressible in general. Hence knowledge of the rules which govern an empirical phenomenon is not sufficient for predicting

its outcomes. In turn, this implies that there is more to understanding phenomena than just learning – even from data alone – such rules.

The case of clinical questions

Big data approaches are very effective in their use to support decisions. However, (in some domains) there are practical problems in their applications coming from data availability to also quality of data. Moreover, if we have data of sufficient detail and quality, does the data capture the relevant real-world heterogeneity? This is a very important question, as it tells us whether e.g., a clinical question can be answered with the available data and with what accuracy. Suppose that we have an Electronic Medical Record (EMR), the treatment analysis, the dataset for clinical outcome prediction, and we devise some classification analysis and the prediction model.

The first obvious thing is that when the dataset brings no relevant information about the predictors of the outcome, it is unlikely that you answer the clinical question.
Moreover, there are several shortcomings in best practices for ML and AI, for sure in the clinical domain. Those concern what follows.

validation (statistical performance)

- Are the reported metrics relevant for the clinical context in which the model will be used? Which corresponds to clinical applicability
- Is the ML/AI algorithm compared to the current best technology and against baselines? This corresponds to the need for benchmarking methods
- Is the gain in statistical performance with the ML/AI algorithm justified in the context of any trade-offs? Which corresponds to a contextual cost-benefit analysis

impact evaluation

- Are the results holding consistency outside the used system? This corresponds to reproducibility, external validity, and generalizability
- What evidence is there that clinicians and patients find the model and its output reasonable? This couples with interpretability and explainability
- How will evidence of real-world model effectiveness in the proposed clinical setting be generated and how will unintended consequences be prevented? This corresponds to feasibility

How to account for variations due to new or unexpected conditions in an open and changing environment, like for instance data non-stationarity?

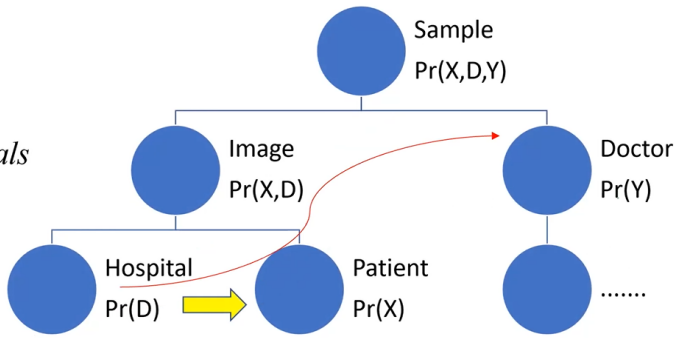
ML/AI algorithms are often trained and validated on huge amounts of historical data. The obtained performance is usually guaranteed only under the assumption that the process generating the data does not change. This is hardly the case in real-world settings, as in practice this assumption is often violated. For instance, in case of medical diagnosis problems, data can come from new clinical sites and then sampled from a different population and collected with different protocols. This causes the ML/AI models to underperform. In general, data distributions always experience a drift phenomenon, causing the ML model to lose predictive power. This depends on endogenous and exogenous factors with respect to the data source.

Moreover, datasets contain inherent biases, linked to the sampled phenomenon. This causes various problems, especially when it comes to integrating data that have their own biases, which will also affect the performance of the classifier as the different biases will sum up. There is also a probabilistic view of dataset bias if one wants to speculate further. For instance, if we think about the medical datasets, data distributions can have specific characteristics, as data from different clinical sites can have different patient demographics, hospitals with different diagnostic procedures, recording equipment, protocols and so on (see the following figure).

Pr(D) is varying by hospitals

Pr(X,Y,D) also varying between hospitals

➔ ***Not independent distributions***



Overall, to consider all the possible sources of variation, we need to consider:

- reaction to unexpected conditions (change, non-stationarity). This brings up the causality versus correlation dilemma, as a causal discovery may ensure more invariance
- technical variation of methods, which recall the reproducibility issue when merging or augmenting data
- phenomenon variability, as in the case of patient's variability and heterogeneity; in this case, multimodal data and information can help as when combined, they can let some properties emerge.

Transfer learning may come into play to aid coping with variability, as a paradigm that is emerging to transfer what has been learnt in one domain into another domain. The problem

with transfer learning is the underlying assumption that the domains of exchange must have solid paths. Depending on the discrepancy between domains, transfer learning can be:

- homogenous, when domains have the same or similar feature space (in this case, only data distribution adaptation is needed)
- heterogeneous, when domains have different feature spaces. In this case, feature space adaptation is needed.

Other approaches are based on Few-Shot Learning (FSL) and meta-learning. The idea behind the FSL is to train a model to recognize several classes for each of which the number of examples available is very small, single (One-Shot Learning) or even null (Zero-Shot Learning). Conventional methods cannot be used to perform this type of training, as the model would immediately go into overfitting. Usually, the FSL problem comes coupled with the Meta-Learning paradigm. The term Meta-Learning indicates a higher level of abstraction with respect to classical Machine Learning because, in this case, the model *learns to learn* through its exposure to multiple learning episodes and uses the experience gained to improve its future learning performance. While in the classical learning approach an algorithm of basic learning (or *internal*) solves a certain task, such as image classification, in Meta-Learning an external algorithm (or *meta-learner*) updates another internal algorithm so that, what the latter learn from the specific task is used to achieve a more complex goal, thus increasing the generalization performance. This paradigm could be more similar to that of a baby that learns a language or recognize a horse.

Another aspect is multimodality, as the more, we mix independent things, the more we gain power in prediction. For instance, in medical diagnosis, each imaging modality performs at different spatial resolutions and voxel dimensions. Moreover, depending on tumor type, imaging modalities complement each other.

In general, one would expect that when a model processes a case that does not resemble any of the cases seen before, it could signal this condition.

At any rate, heterogeneity and multimodality of data is an issue that requires moving away from the assumptions that the learning approaches that has recently experienced this big success will eventually work. A viable approach would be to focus on the so-called *Knowledge Computing*, a paradigm based on *Knowledge Graphs*, which codifies knowledge and puts relations among concepts. Injecting this knowledge into a deep learning model could be a valuable approach. Cross-domain information retrieval and causal reasoning would be supported via this integration. Moreover, it could decrease the burden of training huge deep learning models. Nevertheless, this would require interdisciplinary research, with expertise coming from several disciplines. Moreover, a straight and effective combination of knowledge representation and learning is still under definition.

Can we gather insights into the plasticity and the learning process/phenomenon of deep learning models?

AI systems must provide understandable decisions to support human decision-making, avoiding causing any harm and somewhat embedding ethical values, such as beneficence, non-maleficence, autonomy, justice, and explicability. The latter must be considered a critical necessity, and a way to gather insights into the way AI algorithms take decisions. The field of eXplainable Artificial Intelligence (XAI) enquires about those questions, highlighting open challenges and proposing solutions.

Often, DNNs are black boxes to us because of our lack of understanding of their inner mechanisms. Thus, XAI sheds light on how decisions are taken and provides a definition of what is an explanation. For instance, an explanation can be provided in the input space (e.g., highlighting relevant statements in a text), or through saliency maps, examples, a narrative, or counterfactual explanations. Ideally, an XAI system has several desirable properties, which are informativeness, low cognitive load (easy-to-understand explanation), usability, fidelity, robustness, non-misleading, and offering some degree of interactivity or conversation.

Explainability may be by design (intrinsic explainability) or post-hoc in the case of black-box models. In the former case, the algorithm is transparent by design, thus offering global and model-specific explanations; in the latter case, the explanator is a component to be added to the black box, and it can be global or local, and dependent or not from the model. To understand a decision taken from a black-box model, possible outcomes in the vicinity of the output can be explored to understand what similar decisions look like, and then a decision tree can be built out of such exploration; decision trees are used for this purpose because they are a self-explainable model. The use of variational autoencoders is another technique that can be used to explain the decision taken by a black-box model by exploring (learning) the latent space in proximity of the decision to build a decision tree as an explanator. In the case of intrinsic explainability, semantic properties are used to add transparency to models.

Interpretability of AI outcomes in the health domain

Having interpretable model results is still a big challenge in the health domain. Considering the possible outcomes for patients, the best performance metrics to translate the results provided by AI in a way that they are adherent to those outcomes is the (robust) validation this field is after. Working model proxies are necessary for comparative assessment, considering as minimal requirements:

- the ability to predict the majority class;
- benchmarking against standard statistical approaches;
- the use of the gold standard to check the accuracy.

There is anyway an actual difficulty in judging the results coming from a DNN, beyond the mere ‘convergence to optimality’, because of objective issues in the biological domains, such as:

- lack or insufficient disease-related data;
- poor interpretability of the mechanisms in complex models.

Therefore, the data integration must be contextualized at different levels (i.e., raw input data, features, and predictions) through measures and scores to have a data-feature-prediction mapping. In other words, the performance of an intelligent system should be assessed with metrics linked to measurable outcomes, such as the risk of disease onset, relapse, response to treatment, and so on.

Socio-economic considerations: short notes

Considering the key role that AI is currently playing in decision-support and/or autonomous systems, compliance with legal and ethical principles is mandatory. This means that any AI-enabled system should incorporate ethical values, such as beneficence, non-maleficence, respect for human users' autonomy, justice, and explicability. XAI is a key building block of this standpoint as it strongly contributes to empowering individuals against undesired effects, also ensuring their right of explanations as required by GDPR. In particular, explainers should be devised to respond to questions such as "Am I being treated fairly?", "Can I contest the decision?", "What could I do differently to get a positive outcome?" or "Is my system working as designed?", "Is it compliant?" A key point in this respect relates to the definition of what an explanation is and to how to ensure that an explanation is understandable by users. An ideal explainer should model the user's background.

The development of AI, particularly, DL systems raises ethical concerns also for the efforts and resources needed to ensure high performance. Indeed, thousands of people work poorly paid around the world to pre-process and labelling the data needed to train powerful AI-enabled systems. Another aspect is the right to access and use the data, which must always be clear, following the FAIR approach and avoiding any power unbalance.

Finally, the competition with renowned tech giants could be unfair from several viewpoints. In this respect, the establishment of a large European laboratory on AI, as in the case of the European Molecular Biology Laboratory, would be beneficial to allow Europe to compete at the world level.

References

[1] Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences*, *117*(48), 30033-30038.

[2] “Of course if one waits long enough, the initial state will eventually recur, but the recurrence time is so long that there is no possibility of ever observing it.” Boltzmann, L. (2003). Reply to Zermelo's remarks on the theory of heat. In The kinetic theory of gases: an anthology of classic papers with historical commentary (pp. 392-402).

[3] M. Mezard, *Artificial intelligence and its limits*, Europhysics News 49/5&6, (2018).

[4] H.Hosni, A. Vulpiani, *Forecasting in the light of big data*, Philosophy and tecnology, 31, pages 557–569(2018)

[5] Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE transactions on neural networks, 10(5), 988-999.

[6] Carleo, Giuseppe, et al. "Machine learning and the physical sciences." Reviews of Modern Physics 91.4 (2019): 045002.

[7] Gardner, Elizabeth, and Bernard Derrida. "Three unfinished works on the optimal storage capacity of networks." Journal of Physics A: Mathematical and General 22.12 (1989): 1983.

[8] Y. Jiang, B. Neyshabur, H. Mobahi, D. Krishnan, and S. Bengio, arXiv:1912.02178

Appendix I – background document

CompleXXI and MAD4Future: a joint workshop
The complexity of artificial learning

Introduction and motivations

One of the formidable challenges that science is confronted with in the XXI century is the ambition to deal effectively with *Complexity* in the physical, biological, ecological, and social universe. Stephen Hawking himself declared that this '*will be the century of complexity*'.

Complexity is sometimes ambiguous and is generally accepted that it refers to the emergence of *unexpected collective properties*, a priori unexpected from microscopic interactions. On general grounds, complex systems are characterized by many heterogeneous interacting parts; multiple scales; complicated transition laws; unpredicted emergence; sensitive dependence on initial conditions; path-dependent dynamics; networked hierarchical connectivities; interaction of autonomous agents; self-organisation; non-equilibrium dynamics etc.

From the computing and ICT standpoint, *cyber-physical Systems of Systems* (SoS) are emerging as a paradigm to model the connection, at various scales, of distinct but dynamically interacting systems, whose combination may overall result in complex behaviors. Scientists are thereby challenged to understand how models, algorithms, and data can be leveraged to describe and manage SoS with an organic approach.

Data Science and Artificial Intelligence (AI) are posing themselves as the keystones to face this challenge. The two disciplines are strictly interconnected and are experiencing their momentum due to the Data Deluge that has boosted data-driven learning methods and has enabled the unprecedented understanding of complex phenomena as well as scientific breakthroughs. However, important questions are driving our debate:

- **data abundance is not a synonym of understanding**, but rather calls for a novel synthesis among experimental techniques, high-performance computing, and modeling;
- the **AI computing machineries**, especially those based on the most-promising learning approaches such as Deep Learning, are complex systems themselves and, to date, they do **exhibit several limitations and biases**.

In fact, in the research communities, the great hype around data-powered intelligent machines is being followed by reflections on *the inner limits of learning and data-driven models*. Main concerns pertain to the dimensions of data (quality, quantity, diversity, fairness), and to the fact that obtainable results rely on *statistical correlation instead of causality,* since intelligent machines still lack contextual reasoning capabilities. At last, the yet unproved assumption that, with sufficient data, any complex model could be learnt is clearly emerging. The discussion in this respect is how to deliver of intelligent machineries that can address the limitations of

learning and inferential analyses, including knowledge gaps and convoluted dynamics that depend on:

- context, thus being able to explore causality versus correlation;
- method, thus being able to be integrated into the scientific process;
- variations due to new or unexpected conditions.

On the side of complexity, there is a growing interest in understanding the basic functioning of learning and its intrinsic limits. Theoretical physicists are exporting concepts and methods from statistical mechanics and nonlinear science to grasp the basic functioning principles.

Scope of the workshop

In the next decades, our societies will witness a substantial impact of data-driven and machine-learning methods to sort and extract information. In a world where sensible decisions may be taken based on such inferences, obvious problems and challenges will arise.

Besides the academic interest, a deeper understanding will have an intrinsic social value. For instance, if we are not guaranteed that any deep network will always perform the task for which it was trained and are not aware of its limits, how could we make sensible decisions based on the inferences they provide?

The aim is to enucleate a set of topics where **cross-disciplinary interaction among data, Artificial Intelligence and complexity scientists will be possibly active in the next decades.** In the spirit of the Foresight project, we plan to call representative experts from the two communities to outline the main research objectives.

According to a consolidated approach, the list of participants to the workshops include scientists, policymakers, and industry representatives. A limited number of participants will be asked to introduce the brainstorming among all participants, presenting:

- the possible research pathways with radically new approaches;
- the socio-economic impacts of the proposed approaches, indicating possible funding mechanisms and identifying potential obstacles;
- the challenging aspects, and how novel solutions can be combined with aspects indicating the main obstacles to overcome.

All participants to the workshops are selected to stimulate the debate, achieve a consensus on the new research directions, and prepare a report of the most relevant conclusions of the workshop.

Topics and questions

To guide the workshop, the main topics of tentative discussion are listed in what follows:

- What are the current limitations of inference from data and learning? Is it possible to go beyond them?
- How to use machines to find emergent organizations in our vast data sets?

- Can they be integrated into the scientific process? Can they be used to automate simulations of physical, chemical, or biological processes?
- How to account for variations due to new or unexpected conditions in an open and changing environment, like for instance data non-stationarity?
- Gather insights into the plasticity and the learning process/phenomenon of deep learning models.

Last but not least, it is crucial to find *common languages* among scientists with diverse backgrounds, a task that is by itself difficult and requires close interaction and sharing of knowledge and expertise.

Appendix II - Workshop Program, May 5th 2021

Enrico Capobianco

Lead scientist of Computational Biology & Bioinformatics division of the Center for Computational Science of the University of Miami

Fosca Giannotti

Director of research of computer science at the Institute of Information Science and Technology of the National Research Council, Pisa, Italy

Marc Mézard

Physicist, director of the École Normale Supérieure, Paris, France

Angelo Vulpiani

Professor of Theoretical Physics at Università di Roma "La Sapienza"

Lenka Zdeborová

Professor of physics, computer science, and communication systems at EPFL, Lausanne, Switzerland

Riccardo Zecchina

Professor in Theoretical Physics at Bocconi University, Vodafone Chair in Machine Learning and Data Science

- **Marc Mézard**
  Physicist, director of the École Normale Supérieure, Paris, France
  Presentation title: "Learning and data structure"

- **Lenka Zdeborová**
  Professor of physics, computer science, and communication systems at EPFL, Lausanne, Switzerland
  presentation title: "How many samples are really needed?"

- **Enrico Capobianco**
  Lead scientist of Computational Biology & Bioinformatics division of the Center for Computational Science of the University of Miami
  presentation title: "AI and Machine Learning in Precision and Translational Biomedicine"

- **Angelo Vulpiani**
  Professor of Theoretical Physics at University of Rome "La Sapienza"
  presentation title: "Some Thoughts about Complexity, Data and Models"

- **Fosca Giannotti**
  Director of Research of computer science at the Institute of Information Science and Technology of the National Research Council, Pisa, Italy
  presentation title: "Explainable Machine Learning for Trustworthy AI"

- **Riccardo Zecchina**
  Professor in Theoretical Physics at Bocconi University, Vodafone Chair in Machine Learning and Data Science
  presentation title: "Challenges in contemporary machine learning"

Appendix III - List of participants to the workshops

Ezio	Andreta	CNR Foresight project
F. Manlio	Bacco	CNR-ISTI
Cecilia	Bartolucci	CNR-IC
Enrico	Capobianco	University of Miami
Ruggero	Casacchia	CNR Foresight project
Caterina	Cinti	CNR-ISOF
Sara	Colantonio	CNR-ISTI
Giorgio	Einaudi	CNR Foresight project
Fabrizio	Falchi	CNR-ISTI
Paolo	Ferragina	University of Pisa
Fosca	Giannotti	CNR-ISTI
Stefano	Lepri	CNR-ISC
Roberto	Livi	University of Florence
Marc	Mèzard	ENS Paris
Pier Francesco	Moretti	CNR Unità Relazioni Europee ed Internazionali
Stefano	Ruffo	SISSA, Trieste
Luisa	Tondelli	CNR-ISOF
Angelo	Vulpiani	Sapienza University of Rome
Lenka	Zdeborova	EPFL, Lausanne
Riccardo	Zecchina	Bocconi University