MACHINE LEARNING & CULTURAL HERITAGE

Alfredo M. Ronchi
Politecnico di Milano
alfredo.ronchi@polimi.it

1. Depicting the scenario

To have a clear vision of the "new reality" and its trend, it is necessary to consider the technological evolution together with some relevant events and the impacts that these aspects have had and will have on society, then carrying out a "future back casting" exercise.

Recently, some messages strengthened their impact on society as a mix of incumbent tragedies and an ongoing full reshaping of society, a kind of imminent “new global order”. On the one hand the global warming, climate change, the ozone hole, lack of food and water, the pandemic crisis and more have had a profound impact on society, generating a widespread feeling of risk for the survival of humanity. The “cancel culture” movement, and the negative impact of man on nature are pushing the most radical thinkers of the 20th-century to stop facing the prospect of the actual extinction of Homo Sapiens. This perspective, as the endpoint of the Anthropocene the faith of anti-humanism, begins not with a political program but with a philosophical idea. The flip side, Transhumanism glorifies some of the same things that anti-humanism decries—scientific and technological progress, the supremacy of reason. Some transhumanists believe that genetic engineering and nanotechnology will allow us to alter our brains and bodies so profoundly that we will escape human limitations such as mortality and confinement to a physical body. Others are adamant that general artificial intelligence design will improve itself to think faster and deeper, then the improved version would improve itself, and so on, exponentially. Both trunks of thoughts basically consider humans’ disappearance, on one side extinction on the other cyborg.

There are some key events that have characterised the recent period, one of these is the so-called digital transformation considered the natural evolution of the current society in the light of a pervasive technology like digital technology.

Digital technology is intertwined with almost all the life sectors. Since the dawn of digital technology, the number of application and solutions based on such technology had a surprising rate of growth. Transistors, originally conceived to fight against deafness, were the sparkling light of several new devices, followed by integrated circuits. Computers became ten times smaller and powerful being ten times cheaper. Nowadays there is no field of human knowledge that doesn’t take advantage or is based on digital: communication, education, government, health, energy, mobility, etc.

The full settlement of such pervasive technology is termed digital transformation (DT or DX), governments, international organisation, private companies are all together promoting and facilitating this switch from analog to digital. We are increasing leaving the analog, face to face, paper-based world to enter the intangible digital mediated one. Many years ago, in the 1980s, it was held multidisciplinary panel to discuss about the ontological aspects of digital to approach this sector properly. The outcomes were that digital objects represent a completely new class of objects they enable the opportunity to be here and there, clones perfectly equal to originals, shared among an illimited number of owners, “immortal” in theory and more. It is all gold what it glitters?

2. Taxonomy of ML, ML block boxes certification ID valet

The extensive use of Artificial Intelligence, Machine Learning and Big Data, apart from several ethical issues, can led to some relevant drawbacks. As an example, let’s consider “nudging”. The concept of nudge is already used in digital systems even if the nature of the mechanisms that characterise it is not always consistent, and some uses overflow into practices already prohibited by current legislation. In fact, the use of even "slight" and often morally irrelevant manipulations of the architecture of the decision is constrained both in the use of personal data to be able to construct a nudge mechanism[1] and if the desired result falls within the category of fraudulent transactions[2]. The progress of AI has made it possible to develop much more powerful nudge mechanisms thanks to the effectiveness of statistical and inferential AI systems. The impact of AI powered technology on human autonomy is huge. AI-enhanced nudges reinforce the ability to achieve the designer goals using cognitive biases, emotional impulses, and other human behavioural mechanisms both intentionally and unintentionally.

If on one side the whole architecture is based on cyber tech, with all the potential risks it implies, on the other side cyber-world rules have can express a power that no one of the "rules" in history had, information and big data are the assets to be analysed, influenced, reused. Some authors call them "the new oil" but this type of "oil" can be used, abused, and misused many times. The science fiction "Ingsoc[3]" or "Cyberdyne[4]" now rule thanks to "algorithms" and "neural networks".

3. Artificial Intelligence and ML

Artificial Intelligence (AI), cutting edge technology in the eighties depicted by the press as a dangerous shift of humans towards technological slavery, was looking for a reasonable field of application as it happened in the case of the Japanese stock exchange, unfortunately, some "bugs" in the system generated the crash of the market. The concrete application was addressed to make, among the others, washing machines and camcorders smarter. The traditional domain of Artificial Intelligence, generated along its path some specific domain of application making our software, home appliances, accessories, and cars more "intelligent". This evolution was accompanied by the usual philosophical debate on "Can machine think?". The reference study in this sector is indubitably due to Alan Mathison Turing, mathematician, philosopher, cryptographer and more, and his article "Computing machinery and intelligence"[5] , the first paragraph entitled "The Imitation Game" starts with - "I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the term "machine" and "think." – and then explains his vision on "thinking machines" providing a more sophisticated definition and revolutionary insight on future technologies. Now AI is back on stage with a completely different impact on society.
In the era of open and big data, AI allows extremely large data sets to be analysed computationally to reveal patterns in any kind of datasets (social, political, medical, business, etc), which are used to inform "managers" and enhance decision-making. We used to identify two different branches of AI: "General" also known as "strong AI" and "Narrow" also known as "weak AI".
On one side we find a broad-spectrum artificial intelligence designed to face a wide range of problems "imitating" the human brain, on the side of weak AI, we find vertical solutions based on a well-defined domain of knowledge as it happens for instance for expert systems or car automatic driving systems. They are designed to deal with a specific domain of knowledge, characterised by well-defined rules and situations; they can be further trained and even implement machine learning; additional everyday examples are intelligent personal assistants, chatbots, SIRI, ALEXA, GOOGLE Assistant, Mercedes Benz, and Volkswagen onboard assistants. More in detail:

[1] by the GDPR
[2] UCPD - Unfair commercial practices directive https://ec.europa.eu/info/law/law-topic/consumer-protection-law/unfair-commercial-practices-law/unfair-commercial-practices-directive_en
[3] George Orwell - 1984 big brother
[4] James Cameron's ruling organisation in Terminator
[5] A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460., https://www.csee.umbc.edu/courses/471/papers/turing.pdf last access July 2018.

Narrow AI (ANI) - Narrow AI is a collection of technologies that rely on algorithms and programmatic responses to simulate intelligence, generally with a focus on a specific task. Time ago this was the branch of AI addressed to create expert systems, a software application designed to solve specific problems providing the rationale of the outcomes. When you use a voice-recognition system like Amazon's Alexa to turn on the lights, that's narrow AI in action. Alexa may sound smart, but it doesn't have any advanced understanding of language and can't determine the meaning behind the words you speak. The program simply listens for key sounds in your speech and, when it detects them, follows its programming to execute certain actions. To users, this can seem surprisingly intelligent — and voice recognition is far from a simple computing task — but there is no actual "thinking" going on behind the scenes. Non-player characters (NPCs) in games are another good example of ANI. While they take human-like action, they're simply following a pre-programmed series of actions designed to mimic how a human would play the game.

General Artificial Intelligence (GAI) - GAI, in contrast, is intended to think on its own. The goal of GAI research is to engineer AI that learns in a manner that matches or surpasses human intelligence. GAI is designed to learn and adapt, to make a decision tomorrow that is better than the one it made today. None of this is easy, which is why most examples of AI you'll encounter today are the narrow form. GAI is a new, complex, and varied category with numerous sub-branches, most of which are still research topics in a lab. Modern AI systems focus on solving specific tasks, such as optimization, recommendation or prediction systems and don't learn broad concepts generally, as a human would.

4. Machine Learning

Machine learning (ML) is an interesting subset of AI that is providing inspiring solutions to complex problems, a typical field of application is the one non-approachable with algorithms and explicit programming. The basic principle is to analyse data and identify patterns that can suggest a way to extrapolate a significant result. The typical taxonomy of ML is at the top level subdivided into supervised learning and unsupervised learning.

<u>Supervised learning:</u> a system "tutor" feeds the application with a set of inputs and expected outputs to train the system that has the identify a general rule that maps inputs and outputs; of course, this is a possible option when this "rule" is not clearly identifiable by the software programmer so a specific algorithm it is not doable.

<u>Semi-supervised learning:</u> the system receives only incomplete training, there is not a complete set of outputs related to the list of inputs.

<u>Reinforcement learning:</u> the key feature of this approach consists in a dynamic environment that provides a score (positive or negative) rewarding the strategy to be followed to reach the requested output; thanks to this assessment cycle we can say that the system learns and provide better solutions as much as it runs[6].

<u>Unsupervised learning:</u> the learning algorithm is completely independent, it does not receive any information about the outputs or any score, it must identify by itself the structure of the input and discover potential hidden patterns or identify a potential goal thanks to feature learning.

Supervised machine learning algorithms and models use labelled datasets, beginning with an understanding of how the data is classified, whereas unsupervised models use unlabelled datasets and figure out features and patterns from the data without explicit instructions or pre-existing

[6] Bishop, C. M. (2006), Pattern Recognition and Machine Learning, ISBN 0-387-31073-8, Springer

categorizations. Reinforcement learning, on the other hand, takes a more iterative approach. Instead of being trained with a single data set, the system learns through trial and error and receiving feedback from data analysis. With faster and bigger computation capabilities, ML capabilities have advanced to deep learning, a specific kind of ML that applies algorithms called "artificial neural networks," composed of decision nodes to more accurately train ML systems for supervised, unsupervised and reinforcement learning tasks. Deep learning approaches are becoming more widespread but come with high computation costs and are often harder for humans to interpret because the decision nodes are "hidden" and not exposed to the developer, on the contrary traditional NAI algorithms use to provide the rationale behind the outcomes step by step. Nonetheless, deep learning offers a wealth of possibilities, and already has promising applications for image recognition, self-driving cars, fraud news detection and more.

To better clarify the role of ML we can consider, among the others, two typical tasks it can perform:
<u>Classification:</u> Inputs are divided into two or more classes (labelled); the system must produce a model that assigns additional random inputs to one or more of these classes[7]. As we will see in the following taxonomy this process is usually performed in a supervised manner, the classes are defined a priori. A typical example of classification tasks performed by ML is spam filtering; the two classes are, of course, "spam" and "not spam". The learning process will increasingly add filters to better perform the classification.
<u>Clustering:</u> The task is to divide a set of inputs into groups (unlabelled); it looks like the classification tasks but this time the groups are not known beforehand. This is typically an unsupervised task.

Let's leave this side of the technology to face another relevant one, how to deal with responsibilities in case of accidents that directly involve AI or ML?
If we refer to air control probably one of the closest sectors the choice is usually between technical problems and human factors. Many times, the final verdict is a mix of several causes that all together led to a disaster. Accordingly, with the reports, 70% of aviation accidents can be attributed to human error. Why? Because humans are active players inside the systems, and they are the only components that during emergencies can adapt and adjust resources to try to cope with unexpected events. Of course, these responsibilities are not only in charge to pilots, but they are also shared among organisational failures, conditions of the operators (physical and mental state), physical and technological failures and finally human errors.

We increasingly hear of car driver assistance technologies or even autopilot. In case of law infringement or accident who is in charge as responsible, the "driver", the car builder, the software company, all of them? No one, the fate? As usual in risk analysis in addition to risks due to our behaviour or decisions, we have risks that do not fall under our control. We must consider that even the "road environment" is part of the system, horizontal and vertical signals, timely updates of maps and road works are an integral part of the package. Some lane control systems are cheated by multiple lane lines due to old lines still visible. Some accidents involving "intelligent" cars and even humans already happened, and the responsibilities are not yet undoubtedly assigned. Finally, the ethical aspect will merge with intelligent algorithms.

---

[7] In case of more classes it is termed "multi-label classification".

5. Different approaches to AI/ML regulation

There are several initiatives, apart from national ones, aiming to AI & ML regulation: EU Regulatory framework proposal on artificial intelligence, UNESCO, ITU AI4Good, IEEE, IGF, and WSIS.
On the occasion of the recent WSIS Forum (13-17 March 2023) this topic was approached by different panels and workshops: Can cyber tech be resilient and green?, Responsible AI governance and the challenges of general purpose AI, The Rise of the Machines: Will AI Destroy Your Job?, Global Digital Compact roundtable discussions, Data ethics and the ethics of digital and emerging technologies, Inter-Agency Working Group on AI- Bringing together the UN System expertise on AI, Future of e-government assessment in the era of AI: Opportunities and Challenges, AI-powered Open Educational Resources, Generative AI and the information society: How IT professionals can ensure that generative AI supports the information society, Ethics: Moving from Principles to Practice in Governing AI Using The UNESCO Recommendation on the Ethics of AI, and AI for Good. The outcomes of these meetings contributed to lower the concerns about AI and ML potential drawbacks mainly because of the number and quality of the debates.

Code of Ethics for AI – Seven principles enable to classify the dilemma and overcome:
AI with carefully delimited impact – designed for human benefits, with clearly defined purpose setting out what the solution will deliver, to whom;

Transparent & explainable AI – with outcomes that can be understood, traced and audited, as appropriate;

Controllable AI with clear accountability – enabling humans to make more informed chiches and keep the last say, including feedback plans when needed;

Robust and safe AI – ensure technical robustness of AI from safety, security, and accuracy standpoint;

Sustainable AI – develop mindful of each stakeholder, to benefit the environment and all present and future members of our ecosystem, human and non-human alike, and to address pressing challenges such as climate change;

Fair AI – produced by diverse teams using sound data for unbiased outcomes and the inclusion of all individuals and population groups.

5.1. FAIR AI

Fair AI refers to probabilistic decision support that eliminates or reduces the impact of bias against certain users (Capgemini Research Institute).

- DEMOGRAPHIC PARITY/ STATISTICAL PARITY

The model is fair if the composition of people who are selected by the model matches the group membership percentages of the applicants.

- EQUAL ACCURACY

The percentage of correct classifications (people who should be denied and are denied, and people who should be approved who are approved) should be the same for each group.

5.2. f.AI.ry

Enabled view for enhancing for fairness impact on AI driven business outcomes

EQUAL OPPORTUNITY
Ensure that the proportion of people who should be selected by the model ("positives") & those who are correctly selected by the model is the same for each group.

FAIRNESS THROUGH UNAWARENESS
Remove all group membership information from the dataset. For instance, we can remove gender data to try to make the model fair to different gender groups.

5.3. AI RESPECTFUL OF PRIVACY & DATA PROTECTION

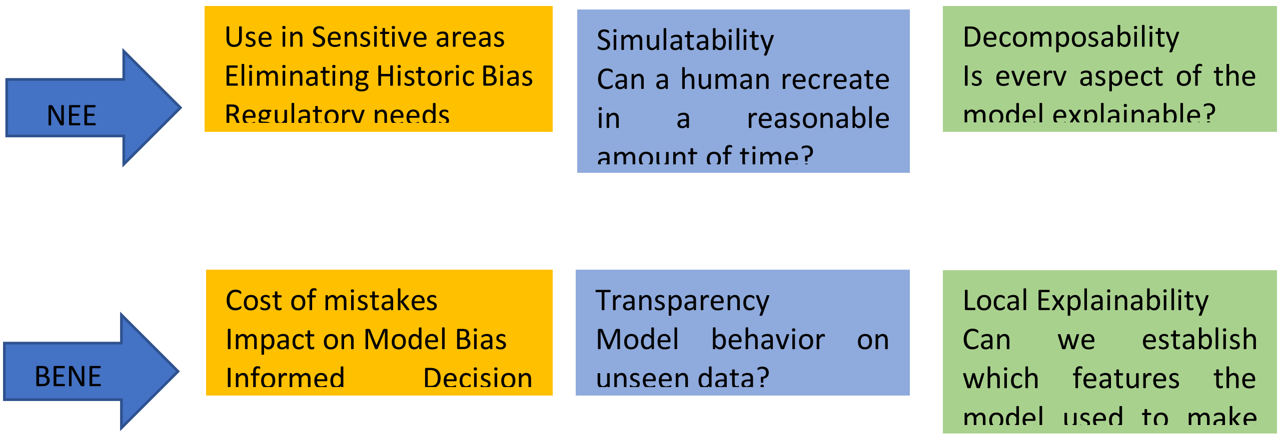
Al teams should incorporate "privacy-by-design" principles in the design and build phase and ensure robustness, repeatability, and auditability of the entire data cycle (raw data, training data, test data, etc.) (Capgemini Research Institute).

Key examples where ai crossed the privacy threshold:
- Clearview Al Faces Recognition
- Facial recognition system scrapes billions of images from across the web used by Law enforcement agencies.

Al Usage for Immoral Intent
- Using Deep Fakes to create chaos by spreading false news.

Practices to ensure privacy and data protection:
- Homomorphic encryption
- Data Anonymization and Data Masking
- Secure multi-party computation
- Federated learning
- Synthetic Data
- Differential Privacy

5.4. Transparent & explainable AI
Explainable Al aims to address how black box decisions of Al systems are made. This area inspects and tries to understand the steps and models involved in making decisions (Forbes).
Al Glass Box - Unboxes black box models and brings maturity to Model Management.



Let's now consider some of the typical big Data and ML tasks like: Data clustering, Data structure, Find patterns / Patterns recognition.

AI works in black boxes; the outcomes are sometimes unpredictable and hard to certify.

6. Machine Learning applications in the field of Cultural Heritage

As already outlined the field of ML into three distinctions – Supervised, Semi-supervised and Unsupervised.
The application of Machine Learning (ML) to Cultural Heritage (CH) has evolved since basic statistical approaches such as Linear Regression to complex Deep Learning models.
How much of this actively improves on the underlying algorithm versus using it within a 'black box' setting. Some key aspects are:

- The lack of suitable training datasets (garbage in and out)
- Classical classification and regression techniques, such as Linear and Logistic regression, have already a distinct and useful application within CH.

The use of Machine Learning (ML) techniques within Cultural Heritage (CH) is still limited, since most of CH literature shows a tendency to rely on statistical toolboxes, which are commonly applied as a 'black box' on small datasets that are not generally publicly available.
Linear and Logistic regression, have a distinct and useful application within CH - While these can be applied in conservation efforts, such as historical building integrity prediction.
Interestingly in the application of Support Vector Machines (SVM) refined the hyper-parameter estimation to support multiple instances learning for recognising iconographic elements in artworks.

Some key aspects are Access to data (CH data repositories), Quality of data (homogeneity, representativeness. granularity, noise, garbage) – no specific privacy issues – mainly text or 2D data few 3D – limited meta-data and link with additional sources, Data collection / interpretation biases, cultural models, The contribution provided by ML is many times not more relevant than the one provided by an expert or a team of experts but in the future …

Typical application of Machine Learning in the field of cultural heritage are Text analysis integration of incomplete text, archaeology, Reintegrate lost part of inscriptions, Reintegrate decoration fragments, Explore big data and extract clusters / similarities /exceptions, Remote sensing (archaeology) Identify potential evidences of manufacts, Analyse big data concerning restoration protocols and identify potential drawbacks, Analyse and identify specific features in artefacts.

Several initiatives are actively working to make AI &ML trustable like to assign a certified digital identity to AI and ML modules and or to add a digital wallet to AI and ML modules to collect certificates. In the meantime, the European Commission is working to issue an AI Act based on risk assessment to frame legally the development of AI and ML.

7. References

[1.] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics) Springer-Verlag, Berlin, Heidelberg (2006)
[2.] G. Carneiro, N.P. da Silva, A. Del Bue, J.P. Costeira, Artistic image classification: an analysis on the printart database, European Conference on Computer Vision, Springer (2012), pp. 143-157

[3.] E. Charalambous, M. Dikomitou-Eliadou, G.M. Milis, G. Mitsis, D.G. Eliades, An experimental design for the classification of archaeological ceramic data from cyprus, and the tracing of inter-class relationships, J. Archaeol. Sci., 7 (2016), pp. 465-471
[4.] L. Chen, J. Chen, Q. Zou, K. Huang, Q. Li, Multi-view feature combination for ancient paintings chronological classification, J. Comput. Cult. Herit., 10 (2) (2017), pp. 7:1-7:15
[5.] K. Crawford, T. Paglen, Excavating ai: the politics of training sets for machine learning (2019).
[6.] A. Elgammal, Y. Kang, M.D. Leeuw, Picasso, matisse, or a fake? Automated analysis of drawings at the stroke level for attribution and authentication, AAAI, AAAI Press (2018), pp. 42-50