Artificial Intelligence Today: Human and Digital Cooperation as the Foundation of the New AI-Based Economy

Rita Cucchiara

University of Modena and Reggio Emilia, Italy

Abstract

This document aims to present and discuss various aspects of the current state of Artificial Intelligence from a technical and practical perspective, with a specific focus on its impact on the world of work, both socially and in the emerging AI-based Economy. The starting and ending point is that the only solution for a future that mitigates risks and maximizes opportunities is not just interaction – now increasingly feasible with generative Artificial Intelligence – but conscious cooperation between humans and artificial intelligent systems for the management of new knowledge. This cooperation, especially, must be directed towards the values of our society, prioritizing research and implementation of reliable, ethically sound, and transparent systems. Human-AI cooperation also requires a new mode of design, public-private interaction, service society, and large Big Tech companies – including international ones – and an awareness that new forms of work are inevitable and will offer great opportunities. Fundamental to all this, however, is the need for regulation, especially at a European and transnational level.

This background paper was prepared in support of the presentation at the HERA strategy meeting: HerAcademy, Bologna, December 5, 2023. The HerAcademy Coordination Committee is thanked for their valuable suggestions and insights for improvement[1].

1. Introduction: towards an *AI-based Economy*

The life of a man or woman in the 21st century is intrinsically digital; fortunately, it is not "only digital." We live and work in an analog world, made up of human relationships, interactions with the environment, speculative moments, and physical activities. Nonetheless, we have learned to coexist with the digital part of the world around us, often referred to by a term, "cyberspace," which now seems outdated but reminds us how the cyber/digital/virtual and physical aspects are intrinsically connected in a new model of reality.

From the digital world, we extract information in its native form (e.g., the digital time on our smartphone) or in processed form (e.g., traffic on a particular road). We entrust our memory, documents, and bank accounts to the digital world. In the digital world, we live to connect with others online, to work or for entertainment. In the digital world, we leave traces of our life: traces that until a few years ago were recovered through searches guided by our knowledge, by explicit metadata (e.g., the date of a document), by queries in structured databases, or at most by simple keywords on web search engines; now, they are traces recovered, correlated, and processed by intelligent digital systems. Indeed, they are now the basis for unsupervised machine learning by humans to train large-scale Artificial Intelligence models.

[1] This document has been written by the author in Italian; then it has been translated through GPT-4 in English. Apologize for possible hallucinations.

We have become accustomed to the need to coexist with digital data: we eat foods whose expiration dates have been calculated, we heal with medicines developed by digital systems, and we work with data that have been digitized and, in most cases, written, stored, assembled, and modified by a computer. If this is true in everyday life, it is even more so at work.

The digitalization of work has led to an inseparable integration of human and digital capabilities, without causing apprehension, prejudice, or fear. On the contrary, public and private investments have multiplied to limit the Digital Divide, to support those who are not able to use digital data in social and productive life. Moreover, Europe has focused on programs for the digitalization of the industry, both manufacturing and services. The Industry 4.0 paradigm, born at the Hannover fair in Germany in 2011 [K. Shlawab, 2015] and then adopted by the Italian government from 2016 as a paradigm of digitalization of the production process, continues to evolve towards extreme dematerialization in virtual and digital worlds by design. Despite the inherent risks of excessive digitalization, both from the point of view of (cyber)security and the loss of specific competence, delegating to others - the large IT Big Tech - the ability to manage and monetize the digital transformation process of knowledge, no one for a decade has shown intolerance, fears, or prejudices about the risks and benefits of the advent of information technology.

The advent of modern Artificial Intelligence has changed the rules of the game.

Complicit in the name that recalls a capacity until now intrinsically human, and an excessive hype by the press and media, never before has Artificial Intelligence (AI from now on using the English acronym) been at the center of debates not only technological and economic. The fear that machine learning from data (Machine Learning) is out of control of the domain expert, that the person at work loses their authority, and that the power of the new AI technologies is superior to human management has led to prejudices and sometimes even to the rejection of innovative solutions.

This is not to say that the debate is useless and that the fears of excessive, uncontrolled consequences of the latest AI solutions are entirely unfounded. In fact, what is happening with the proposals and discussions on the European regulation of the AI Act from 2021 [EU AI Act, 2021] and the recent international summits such as the Hiroshima Declaration [G7 Hiroshima, 2023] of the G7 and the Bletchley Park Declaration [Bletchley, 2023] of November 3, 2023, of 27 countries at the AI Summit, is the need to understand and regulate the social and political effects of a very powerful technology, largely in the hands of a few and used by many. However, often the debates do not have a technological foundation, as they should: the scientific-technological peculiarity that has led to the results of AI, both small and large scale, and to the development of modern Generative AI and Large Language Model (LLM) conversational models like the famous ChatGPT, is only rarely central in discussions that risk remaining in the abstract realm, analyzing the effects of an "intelligent" machine, as if it were a new biological species and not - as it indeed is - the fruit of human science and engineering.

For this reason, the main objective of this document is to present the opportunities and limitations of current AI in its impact on work under an engineering light and as objectively as possible, to understand the dangers but above all to emphasize the potential of emerging technologies.

Information technologies have always sparked debate. In truth, we have witnessed several discussions over the last decade about the potential danger of the spread of digital tools, from their origins to the evolution of various informational milestones (Fig.1):

- Some apprehensions about the excessive power of the Web and Semantic Web (without much outcry over the monopoly, since around 2010, of Google's search engine; Google has outperformed the competition, leading to the closure of its major competitor Yahoo in 2021 and now managing more than 80% of the market compared to competitors such as Russia's Yandex, or Microsoft's Bing, even despite the integration of conversational tools like ChatGPT on Bing);

- Some fears about the excessive use of mobile phones and apps (more for the presumed danger of the devices than for the inherent risk of data sharing) that for more than fifteen years have been indispensable in work and daily life;
- Some hesitations about the use of cloud computing (a paradigm that has shifted our data from a "secure" home server and backup to storage in a remote and initially unknown location);
- Long debates on the pros and cons of social networks (Facebook, Instagram, TikTok, X, just to name a few), as a tool for aggregation and an excellent economic tool for e-commerce and at the same time a tool of power for data owners and companies capable of spreading true and knowingly false news (a debate probably late after the realization of the excessive influence of fake news, intended by humans and not by artificial systems, moreover);
- Various discussions on Big Data, from the first conceptualizations by IBM in 2012 [IBM, The FourV 2012] also thanks to the simultaneous spread of devices in the "Internet of Things" [O. Hersent et al. 2012] greeted with enthusiasm at the base of digitalization, but overshadowed by problems concerning the ownership and sharing of Data, a discussion that has now culminated in Europe in the Data Act of 2021.

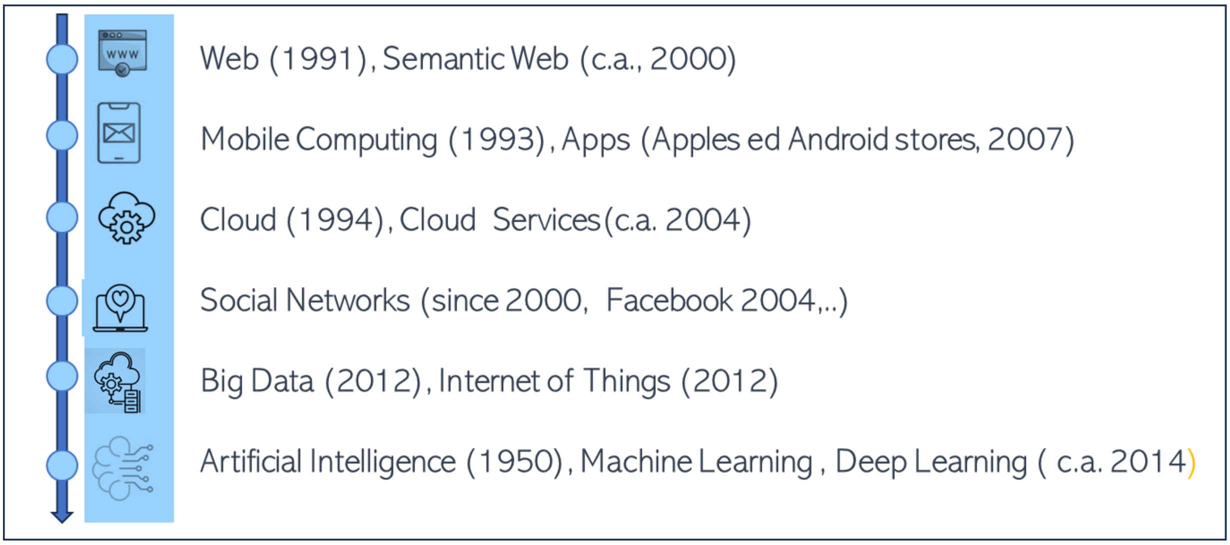

Fig.1 shows the milestones of information technologies for digitalization: *from the birth of the Web at CERN in 1991, the Semantic Web from the 2000s, to today (www.Statista.it 2023), where access to the Web is shared with about 5.4 billion people. Similarly, apps (abbreviation for applications) are the products of Mobile Computing whose initial theories date back to 1993, but from the first app stores in 2007 to now (Forbes, 2022), more than 37 million apps have been counted, of which about 28 million are on Android. The concept of the Cloud was introduced by D. Hoffman in 1994, but the first virtualization services of Amazon Web Services are from 2002, and now they are an essential tool for work. Up to AI, whose first theories derive from A. Turing in 1950, but which have now materialized in the first results of Deep Learning, at least since 2014.*

The debates have generally been constructive: they have not limited the use of technology but have instead strengthened its capabilities and beneficial aspects in digitalization and conscious use. Indeed, the digital evolution of the millennium has marked the transition from a knowledge economy to a data-driven economy [D. Ciurial, 2017].

While the *Knowledge-based Economy* paradigm spurred the development of intellectual supremacy, rewarding companies capable of capitalizing and protecting their strategic-productive know-how, intrinsically based on people and their knowledge, the paradigm of the second decade of the century shifted to a *Data-driven Economy*, delegating productive and economic power to companies capable of owning and managing data: not only the big IT companies of the Web (Google, Microsoft, Amazon, Baidu) that own individual data, but also companies capable of aggregating and managing the data of smart cities, economic phenomena, and companies able to protect their corporate data.

We must now coin a new term for the society of this third decade of the century, which will increasingly be based on an *AI-based Economy*, whose power materializes in a new form of knowledge, learned from data albeit controlled by human expertise.

The new AI-based Economy is the main theme of this document.

It aims to emphasize that current AI systems are designed to generate a new form of knowledge, often not explicit and not necessarily immediately interpretable by humans, but much deeper, capable of being the basis of multiple inferential processes, generalizing in different fields, and whose use significantly modifies every social and work activity. The knowledge, technically indicated as constructed in the latent space of neural networks, built through phases of machine learning, is the foundational aspect of all models and tools of AI based on Deep Learning and the new Generative AI, capable of generating data as well as knowledge from its latent knowledge.

Tools and products based on these paradigms of Generative AI are transforming society, work, and the economy in an irreversible process. This document intends to discuss their characteristic aspects, mainly discussing the positive impact that all agree on, but will also briefly address the challenges, risks, and difficulties entailed by too rapid development and often unconscious use of Artificial Intelligence systems.

The document also aims to provide basic literacy on Artificial Intelligence, although for non-technical use, to lay the groundwork for an informed discussion on its utility and expected impact on the world of work. In the following pages, philosophical or abstract discussions will be avoided to focus as much as possible on what current technology can allow and how the concept of work is changing, necessarily in cooperation between natural and digital intelligences.

A note on the title: "*Artificial Intelligence Today: Human and Digital Cooperation at the Foundation of the New AI-based Economy*" is a manifesto to the awareness that society and the economy cannot afford to talk about AI in the future. We are dealing with the present, where only a conscious collaboration between human and digital skills and intelligences constitutes the foundation of the AI-based Economy in which Italy must play a leading role.

2. Artificial intelligence a technology in a rapid evolution

Artificial Intelligence is simply the latest information technology to reach a sufficient level of maturity to be employed as a digital tool in work and everyday life. AI is a remarkably effective software technology, capable of making machines "intelligent" in understanding, generating knowledge, and interacting with the environment and humans to undertake actions.

The definitions of AI, Machine Learning, Deep Learning are abundant in the media, but the definitions we use in specialized university courses (like the new Master's Degree "AI Engineering" at the University of Modena and Reggio Emilia) are those defined by the European Community in "AI for Europe" 2018 [EU AI for Europe, 2018] and in official documents such as the OECD's "White paper on AI" 2019, updated in November 2023 [OECD Legal, 2023] (Fig.2).

Beyond technical definitions, an operational definition can simply be that "*AI is the intelligence demonstrated by machines as opposed to the natural intelligence shown by humans and other animals.*" This definition was provided by an AI system (BARD, the Large Language Model by Google, 2023), trained on billions of documents written by humans. It is a strong, albeit tautological, definition with which I agree, affirming a now definitive fact, namely that the landscape of "intelligences" has evolved and that it no longer even makes sense to debate the meaning of the term or its performance comparison with natural intelligence.

In many areas (such as facial recognition, medical diagnosis, responding to historical or legal queries requiring access to large sources, weather forecasting), artificial performances are even greater than those of human experts or software not based on Machine Learning. On this premise, it is clear the need to find, in everyday life but especially in work, effective modes of cooperation with humans and regulation.

Artificial intelligence (AI) refers to systems that exhibit intelligent behavior by analyzing the environment and making decisions – for specific objectives and with a certain degree of autonomy..

«AI in Europe», UE Commission2018

- AI: "Artificial intelligence systems are machine-based systems that can, for a given set of human-defined objectives, make predictions, recommendations, or take decisions that affect real or virtual environments.
- Machine Learning (Automatic learning) [...] a set of techniques that allow machines to learn automatically through models and inferences rather than through explicit instructions from a human.
- Neural Networks: Neural networks involve the repeated interconnection of thousands or millions of simple transformations in a larger statistical system that can learn sophisticated relationships between input and output. In other words, neural networks modify their own code (the parameters or weights) to find and optimize connections between input and output."
- Deep Learning (deep learning) is a paradigm for changing the representation of data into a highly compact representation at a higher level of abstraction, depending on the target objective, using various deep layers (deep) of neural networks."

"White paper on AI" OECD 2019

***Fig.2 outlines the definitions of AI, Machine Learning, Neural Networks, and Deep Learning** as derived from the European Commission 2018 and OECD 2019-2023. The definition of AI, both European and from the OECD, reminds us how intelligent systems can use inputs received from the environment (text, images, sounds, sensory values, etc.) to generate predictions, decisions, but also content; for the EU definition, it's important to emphasize the degree - limited - of autonomy.*

AI, through Deep Learning tools, can now work with similar methods on different problems and handle multimodal and diverse data such as images and videos, text, audio and voice signals, sensory data, time series, tabular and digital data, ultrasound and biological signals. The power of the latest neural models, especially the large-scale ones called "*foundational models*" [R. Bommasani et al., 2021] for their nature of defining a latent knowledge that serves as a foundation for specific applications, lies in their ability to transform different inputs (image pixels, numerical series, text, etc.) into a compressed digital vector space (the latent or embedded space) where it is easy to extract correlations, classifications, and concept recognitions or generate new data and knowledge. For this reason, in Machine Learning, we have been talking for many decades about discriminative models (e.g., to classify a benign tumor different from melanoma) and generative models (e.g., to generate a text or an anomaly prediction). Generative AI, despite its ancient modeling, has only in the last few years refined neural architectures to be effective, succeeding in generating long texts, conversations, songs, or images. It is studied at the research level, but many products are already on the market such as the

GPT family, PalM, LLama, Dall-E, Stable Diffusion, and many others, also of smaller size and designed for specific application domains.

Some of these Large-Scale models are on the market with "free" access in the basic version and then for a fee in the enterprise version or as APIs to be used for the generation of new products. An example is the line of AI systems produced by OpenAI, which, after producing GPT and ChatGPT (version 3.5 of GPT) in 2022, now has GPT4 with Dall-E2 for image generation on the enterprise market and is opening a GPTStore for systems dedicated to specific conversational tasks. Others, like META's LLama, are inherently open-source.

AI is now a vast discipline that shares with other information technologies aspects such as ubiquity and pervasiveness but is surpassing all others in its rapid evolution.

Why such a rapid evolution?

Its most striking results are still in the realm of research, but never before in this field has the time-to-market from research to product been almost nullified, allowing for immense dissemination, even at the expense of the stability of the results. This phenomenon is quite rare in other technological areas: consider the very long times between scientific results in biomaterials and their use or between technically already functioning Autonomous Driving systems, out of the research domain for almost a lustrum but still waiting for mass diffusion. The reasons for such speed in the evolution of Deep Learning systems and their transfer from research to product are manifold, but we can condense them into three points:

1. *The democratization of AI scientific research*: AI scientific research is open and internationally collaborative (even in these times of political friction between nations). In the most important AI scientific conferences, explicit descriptions of reproducibility of procedures are required, indicating open datasets and available software code (often on sites like GitHub). In recent years in the AI scientific sector, conference papers have become more authoritative (especially in conferences like CVPR, NeurIPS, ACL, to name a few) and sought after than journal papers, which require too long gestation times; this has greatly increased the dissemination of knowledge, the ability to cite other works and use them as references for new proposals, and climb on the shoulders of giants even by young researchers and Ph.D.s from around the world. To such conferences – like modern technological gyms – now also participate large companies and SMEs that directly learn and develop AI. On the downside, the democratization of research has led to a massive spread of prototypical systems that are not sufficiently safe or tested but are available open-source, which are then adopted by companies with little internal expertise and by novices, with unsatisfactory or sometimes uncontrolled results.
2. *The close collaboration between research centers and IT companies*: at this moment the major research centers (also in terms of publications) are private centers such as Google's DeepMind or the laboratories of Meta, Nvidia, or Amazon. Furthermore, IT companies, both Big Tech and medium-sized companies, and even start-ups, constantly collaborate with public research centers and universities, funding Ph.D. students and supporting open-source research with significant investments. Many systems become closed and sold products in their industrialized version (like GPT4 in interaction systems, or Mobileye's models for autonomous driving), but only after long interactions with research. Fortunately, this also happens in Italy, albeit on a much more limited scale than in the United States. For example, the Aimagelab laboratories at the University of Modena and Reggio Emilia have been supported for many years by major IT companies like Facebook or Nvidia, and by companies that allow collaborative research in application domains, such as Tetrapak, Leonardo, Prometeia, Florim (for details, see Aimagelab.unimore.it).
3. *The absence of unified AI regulation* has certainly allowed for limitless evolution in AI research and production. However, this lack of regulation has also led to significant discrepancies, privacy management issues, and often the crossing of ethical boundaries – not uniformly recognized in all countries. It has also led to collective experiments, such as those seen in recent months with the collective and unconscious training of OpenAI's ChatGPT using an approach called RLFH (reinforcement learning from human

feedback), employing all human responses to learn and calibrate better, asking directly from users. The absence of regulatory limits has led to the development of highly advanced intelligence models, some of which have entered the market, but many are still available in open-source modes. This status quo is no longer acceptable as potential risks tend to outweigh the undeniable benefits, leading to a global consensus on the need for regulation.

These reasons, along with a genuine market and societal interest, have enabled unprecedented AI market development. Many predictions about the explosion of AI have been known for years and have been confirmed in reality. From the European Community documents of 2018-2020 to market analyses, there is agreement on the exponential development of the AI market. According to 2020 forecasts by PWC [PWC, 2019], "AI could contribute up to $15.7 trillion to the global economy by 2030, of which $6.6 trillion will be due to increased productivity and about $9.1 trillion will be related to new products and aspects linked to AI consumption." These forecasts have improved post-COVID, particularly due to the significant interest in the medicine and health market.

For economic analyses on the AI market and various topics, refer to economic documents such as those from Fortune Global Market [Fortune, 2023] and McKinsey Generative AI [McKinsey, 2023].

3. The directions of Artificial intelligence and the impact in the labor market

AI in 2023 is heavily based on Machine Learning and, in particular, Deep Learning. This involves automatic learning from data through computational models known as deep neural network architectures.

As depicted in Fig. 2, this terminology stems from the fact that data undergo numerous transformations across many ("Deep") computational layers (dozens or hundreds), each consisting of various linear and non-linear operations. In the linear operations, often in matrix form, data are multiplied by "parameters" or "weights," reminiscent of the weights of synapses in natural neurons. In the non-linear operations, certain thresholds (also parametric) are applied, allowing the computational layers, almost like neural layers, to activate if the result is above a threshold and thus transfer information to subsequent layers. The Learning phase, which initially occurs before the networks are employed in the working phase (often called the inference phase), is used to define the value of these parameters. The parameters, initially initialized between values 0 and 1, are "learned," or refined, in the training process of the network, through an optimization process depending on a cost function (or loss function, the "loss"). The choice of this function, along with the network structure and learning methods (the so-called "hyper-parameters"), represents the know-how of the AI Engineer who defines what the network should learn, usually in cooperation with domain experts.

Neural networks are not a recent discovery: experiments with dozens of neurons for basic problems were already conducted in the late '80s. The revolution of about fifteen years ago is attributable to the definition of new architectural models (the "Convolutional" networks) initially with a few thousand neurons, and then from 2017 the architectures defined as "Transformers" [Vaswani et al. 2017], which have enabled the creation of Generative AI and the development of networks with millions of nodes and billions of parameters.

Considering the research, custom solutions, and commercial services of Artificial Intelligence, we can outline three complementary directions: a) AI for data analysis and generation, b) AI for autonomous systems, and c) AI for better AI. These translate into differentiated technological efforts, albeit partially overlapping, that impact the world of work in different ways, implying opportunities and risks.

3.1 L' AI for data analysis and generation

AI for data analysis refers to systems designed to aid humans by providing better tools for data interpretation through direct response or interaction with those who should manage or use the data. AI systems are designed to understand data (e.g., for classification, identification, recognizing specific patterns in data) and generate additional information and data themselves. Traditionally, in Machine Learning, the former are called discriminative models, and the latter generative models: both deal with data to ideally communicate with humans or provide specific information. The term "ideally" is deliberate: systems are now being designed where artificial systems interact with other artificial systems to generate complex systems (e.g., in manufacturing where AI systems can generate sequences of commands to move robots, or to correlate data from other platforms). However, the ultimate goal remains to provide better information for use by a person, user, or professional domain expert. Examples include medical data analysis systems for diagnosis or prognosis [Stefanini et al., 2022], financial data correlation systems for banking operators [S. Luetto et al., 2023], or intellectual interaction through text and images, artificial vision systems used for facial recognition for biometrics or interaction in Automotive [A.Palazzi et al., 2018].

This is the most well-known area of AI, where major efforts of Big Tech are also concentrated, often both owners and producers of large amounts of data, interested in understanding the data and generating data and knowledge. The models provided as closed products, libraries, and APIs, or as open-source systems are generally intended for use in the cloud, with increasingly shared data, sometimes requiring specific training tools (e.g., Federate Learning models, an emerging model to ensure data privacy). Integration Fusion models, which combine the penetration of diverse data and solutions aimed at explainability, become very important, such as in the medical field [Holzinger et al., 2022].

Machine Learning thrives on data, and the recent trend is increasingly towards pre-trained neural architectures on billions of generic data, processed often for days and months on GPU servers. Large-scale models are generative, designed to generate information of the same or different nature as the source data. Technologically, the currently used model is the Transformer (left in Fig.3), which has revolutionized research and commercial products for six years: it is the basis of conversational and reasoning systems, including ChatGPT (Generative Pre-trained Transformer) by OpenAI in collaboration with Microsoft. For example, GPT3, one of the first large-scale systems and the basis of ChatGPT (initially in version GPT3.5 and currently GPT4), had 175 billion parameters in 2022 and was trained with 45 Terabytes of documents in a self-supervised manner, i.e., without the need for specific human annotations. The number of neural parameters and the size of the training data grow together as systems become more complex. In addition to pre-trained models, many others, especially in industrial or medical fields, are trained in a supervised or self-supervised manner on specific data. For example, Deep Learning solutions have been trained to generate large ceramic surfaces with ever-changing patterns from small examples provided by designers [R.Cucchiara et al., 2022]. Pre-trained models can be used as standalone models (now widely used in the workplace as a query tool and, more importantly, as a co-pilot for software code writing) or refined through Domain Adaptation models to be more suitable for specific domains like financial predictions, generating pharmaceutical 3D models, supporting designers, or predicting time series like energy consumption.

The current strength of AI lies in its capacity for generalization, making it feasible to employ similar technologies in vastly different data analysis contexts, making AI seem like a simple science and technology – while it actually requires highly qualified computer expertise – but is certainly of direct use in various knowledge domains. This is why AI models, especially generative ones, are spreading in the world of work, both in support of production and intellectual work and Robot Process Automation, for back-end IT activities like systems for summarization (generating summaries of long and complex documents), interaction and Q&A (Question & Answering), and generating texts and visual or multidimensional material, even in support of creative work

.

***Fig. 3 illustrates the current ingredients of large-scale AI:** generative architectural models (on the left, the Transformer, introduced by DeepMind in 2017); in the center, multimodal data; on the right, examples of GPU servers or HPC systems like the Leonardo in Bologna, 2023.*

The field of data analytics through AI with pre-trained models is where problematic aspects and the need for regulation are most keenly felt, especially due to opacity in the learning phase (what data were used? Copyrighted data, prejudicial or dangerous data?) and at the same time for the intellectual power and interaction capability. Recent technology has become very refined: while the results of Large Language Models in the linguistic field are striking, in the visual domain it is still improving, although tools like DALL-E 3, associated with ChatGPT, or Stability AI's Stable Diffusion, now provide visually impactful images.

However, it's crucial to remember that the results of generative systems are not – and cannot be due to their statistical nature – perfect. Generative systems suffer from a problem metaphorically called "*hallucination*," where texts, sounds, or images that are perceptually correct but semantically wrong or even outright false are generated. It's important to remember – a point not intuitive for non-experts – that generative systems create new data based on probabilistic models: the possibility of generating data that does not correspond to reality, that is, different from the data used in learning and nonetheless implausible, is always possible, especially when complex or out-of-the-ordinary results are requested (often referred to in AI as "long-tail", the long tail of the probability distribution).

Therefore, human validation is necessary. In the field of software production, generative systems are spreading reliably, partly because they are aimed at competent interlocutors like computer engineers or programmers expert in the field; they are well accepted in the healthcare and medical field, where medical experts are accustomed to using technologies in support, for diagnostics, for example. They can also have a very positive impact in managerial, economic, forensic work, provided that AI expertise can be combined with that of domain experts. It will be increasingly important for companies intending to adopt AI solutions to equip themselves with validation tools for processes and risk management, capable of evaluating business choices and ensuring the reliability of AI systems used in decision-making processes. In the near future, risk management systems may become mandatory under European legislation for "high-risk" use (using the current designation of the AI Act), such as in HR recruitment, health, finance, education. It's hoped that risk management and validation systems can be employed throughout the production chain that adopts Artificial Intelligence, to maintain awareness of AI's limits and the transparency of solutions.

3.2 AI for autonomous systems

Many AI solutions and systems are designed and realized as tools to "embed intelligence in machines," equipping autonomous and semi-autonomous systems with the ability to comprehend, decide, and act. This domain has developed over time (Machine Learning systems supporting robotic visual analysis have been around for 20 years) and will continue to evolve in the future to create intelligent vehicles and transportation systems, autonomous drones, autonomous robots in both industrial and assistive environments, as well as smart appliances and objects. The escalation in the last two or three years in the development of autonomous robots, including anthropomorphic ones (like Tesla's Optimus introduced in 2022), combining the best mechanical and materials technology with new cognitive skills, is creating an irreversible trend in research and market adoption. Alongside extreme systems like the next-generation robots, there are already many machines on the market equipped with AI (even floor-cleaning appliances or lawnmowers use vision systems and AI), and partially autonomous systems that still have many intelligent tools on board.

AI for autonomous systems is intended primarily for machine use and shares solutions with the previously described section but also has some peculiarities. For example, there's a greater focus on realization and consumption, related to the cost of production and maintenance, and more oriented towards embedded systems (created with dedicated hardware) than cloud solutions as in the previous case.

According to the European definition, AI should be geared towards systems that take action with a certain degree of autonomy, and one of the premises of responsible AI models (also according to European definitions) is the postulate of "AI oversight," i.e., human supervision. Intelligence is designed by human experts, and even when aimed at the arena of the industrial world and autonomous systems, it must be designed with safety, reliability, and supervision that require human control.

The impact on work by autonomous systems is more silent but definitely of great value. Products in this domain are growing more than linearly, both in the area of sensors and objects in the Internet of Things (according to statistics, there are 16.7 billion connected IoT objects in 2023, expected to rise to 29.7 billion by 2027 [IOTAnalytics, 2023]), and in drones, whose market from $29 million in 2021 is exploding to $63.6 billion in 2025 [Insiderintelligence, 2023] for commerce, internal logistics, surveillance, and many other areas. Added to these are the markets for autonomous and semi-autonomous cars, industrial robots, domestic robots, and many other mobile objects. Clearly, this has an impact on production aiding the manufacturing world in both B2B and B2C, but it also impacts work: they will replace manual activities but will require new competencies for management.

3.3 AI for better AI

Increasingly in scientific research and AI products used extensively as tools to replace or assist work, there's a need for "better" systems. AI is still quite improvable in many performance aspects. Most of the debates we witness daily in the media stem from awareness, or sometimes fear and prejudice, about the fallibility of artificial systems and their failure to align with the tenets of society. Concurrently, there's increasing discussion about uncontrolled AI that could generate systems risky for humanity. Beyond easy alarmist positions, emphasis should be placed on the fact that the international community, especially the scientific community, wants to emphasize potential risks. For a calm but truthful discussion on the subject, reference is suggested to recent interviews with Geoffrey Hinton, one of the fathers of AI, Turing Award winner in 2019 for the invention of Deep Learning [G. Hinton, 2023].

7 Requirements for a Trustworthy AI
• **Agency and human oversight**
• **Robustness and safety**
• **Privacy and data givernance**
• **Transparency**
• **Diversity , non discrimination and fairness**
• **Social and enviormental welness**
• **Aaccountability**
(EU Commission 2020)

Table 1: European Requirements for Trustworthy AI

In this third dimension, AI's direction is not so much towards the individual or machine but towards society, even though it ultimately reflects on the individual, the environment, and production itself for the development of ever better, more reliable, and trustworthy systems. This is particularly vital when systems aim to influence critical areas such as health, environment, or energy resources.

Trustworthy AI has been a European mantra in AI systems since the time of the White Paper in 2020 [EU, White Paper 2020], which defines the 7 requirements shown in Table 1. It's noteworthy that these requirements are now adopted by almost all nations, including Italy, though sometimes expressed differently as "Responsible AI." This term refers not only to systems correctly designed according to ethical models but also to the conscious use of technology.

The term "transparency" in the European formula encompasses both the concept of recognizability (to recognize an AI-generated response or an image produced as a fake from a natural one) and interpretability or explainability to make the rationale behind the provided response understandable. This characteristic of artificial systems being somewhat opaque and needing improvement in terms of human interpretability is a significant aspect in safety-critical systems and those related to health [Holzinger et al. 2022]. AI systems based on machine learning are often referred to as black boxes. In reality, they are not black boxes but highly complex ones that can be understood only through similarly complex evaluation processes.

The problem of transparency is keenly felt, even from the standpoint of understanding the artificial or human nature of a digital product. Therefore, efforts are multiplying to discuss "fake detection" solutions to distinguish the results of artificial systems from real ones. It is hoped that generative systems will increasingly be able to directly insert information about the generative process, which is possible through solutions of automatic "watermarking," or labeling (even if not visible to users) that ensures the data derives from AI and is not a natural product. However, the boundary is quite blurred. In fact, all systems that use automatic translations are now generated by AI; systems that generate completely artificial images are logically very similar to those prepared by human experts, perhaps using AI-based co-design tools (Adobe Photoshop is a striking example). In the fields of creativity, fashion, and architecture, transparent generative systems or not are increasingly used. In a recent interview, Patrik Schumacher, the leading architect of Zaha Hadid Architecture (ZHA), one of the most famous architectural studios in the world, declared himself a great supporter of generative AI as a co-design tool. And finally, generative systems like ChatGPT, which allow textual responses to questions or compositions of even very sophisticated documents, produce results that are difficult to distinguish from human products, both for syntactic correctness, lexical richness, and semantic coherence. The lack of transparency thus pairs with an inherent inability to assess the quality and veracity of the results. These are among the most debated issues in the field of scientific research in AI for the new generations of more reliable and understandable products.

The aspects mentioned above go beyond philosophical digressions but must also be combined according to their technical aspect. Machine Learning and AI systems are still in their technological infancy; they have undergone rapid and uncontrolled development, often are not ready for the market, and the results do not achieve the degree of safety and ethics expected. These are engineering flaws or excessive marketing risks, and thus the ethics of the outcome should be attributed more to the design, development, and production phase (by humans) than to Artificial Intelligence itself. For example, the requirement for non-discrimination and fairness translates into the desire/hope to create systems free from "bias."

Technically, the bias in AI systems (often seen as prejudice, humanizing technological systems too much) is in many cases nothing more than an error or poor engineering design where systems are trained with unbalanced, incorrect, or dirty data. In several situations, this arises for economic reasons, as data cleaning and management for learning is an expensive process requiring costly human intervention. In some cases, this is inevitable (for example, in training with millions of documents in foundation models); in others, it may even be maliciously sought, reminding us that training is part of the design (or domain adaptation) and can be absolutely defined for criminal purposes considering that – even if it can be considered an oxymoron – Intelligence is not artificial [R. Cucchiara, 2021], and even AI systems are designed by humans with not always positive intentions.

If it's true that many AI systems are flawed because they are unconsciously trained with unbalanced data (a real problem in the industrial environment, e.g., in anomaly detection), in some cases, fallibility is accepted when systems are trained with data representing unvalidated statistical history, allowing semantically incorrect or unfair correlations. A simple but important example is if decisions on the remuneration of a female professional were only tied to past statistics, they would forever be lower than the remuneration of corresponding male professionals. Especially in Italy. However, it is realized that the possibility of completely eliminating bias is a lost cause, especially since especially large-scale systems are trained on such large amounts of data that they cannot be verified a priori.

To the problems of bias with generic data are thus added all aspects related to the ethical concepts of our civilization, the right to privacy, intellectual property, human rights, including gender equality in work and society, and the commitment not to process and generate inappropriate, toxic, or violent content. These issues are not yet completely resolved, despite large-scale commercial systems incorporating "Safety check" systems both at the entry point (to filter out violent or inappropriate material) and before output, prior to generation. This is why, for instance, ChatGPT currently refuses to provide responses to inappropriate queries (e.g., on pedophilia or terrorism) and StableDiffusion produces a black image if controlled by an inappropriate textual prompt (input). Unfortunately, safety checkers do not function at 100% and can be bypassed. This necessitates new solutions, which many research laboratories are exploring. One example is the research being done in Modena on "*unlearning*," or the ability to make networks "forget" concepts or redirect their implicit knowledge to "safe" areas of the latent space, to prevent the by-default generation of toxic content. These are only the initial experiments conducted within the European Project ELIAS (European Lighthouse for AI sustainability) of Horizon Europe, [ELIAS, 2023], but they are very promising towards solutions that can make the content generated by generative AI networks less inappropriate.

4. Artificial Intelligence: A Tool for Cooperation with Human Work

Considering the developments in AI and the challenges yet to be resolved, a legitimate question concerns the current use of AI in the world of work. This question is at the heart of ongoing debate among skeptics and enthusiasts and is also the focus of international commissions and government panels, including the XI Commission "Public and Private Work" of the Chamber of Deputies and the "Expert Coordination Committee for the AI Strategy" within the AGID (Agency for Digital Italy) at the Presidency of the Council of Ministers, whose work began in November 2023.

It is now evident that AI in the world of work can offer great potential, although risks and difficulties should not be underestimated. According to a 2022 OECD analysis interviewing 7,300 employees and workers, mainly in the manufacturing and financial sectors [OECD, 2023], 63% of respondents reported being satisfied with the introduction of AI tools in the workplace, eliminating repetitive and sometimes dangerous tasks and allowing them to focus on more complex issues. For the time being, considering the enabling aspects, we can divide the impact of AI in the world of work into two categories:

a) *The adoption and development of AI in various productive domains*, specifically supporting design and production;

b) *The adoption and development of AI in intellectual, organizational, and professional work* to support the management of organizations (risk analysis, HR, management, service operation, supply chain management, etc.) in every public and private institution.

In these two areas, where again the employed AI technologies are quite similar, the common feature is the necessity for interaction between natural and artificial intelligences and, even more so, the need for cooperation.

Interaction with machines and computers now relies on Artificial Intelligence.

The interaction with digital systems is not new, indeed, it's a fundamental prerequisite for the usability and effectiveness of digital systems. This is why, for more than 30 years, computing has specifically dealt with human-machine interaction (HMI), a term used primarily in industrial and manufacturing contexts, referring also to industrial robots. Research in the last decade, with the paradigms of industry 4.0 and the incorporation of AI-based capabilities into industrial robots, has led to the concept of "collaborative robots," i.e., robots capable of collaborating with the operator, and at least equipped with safety systems to avoid harming the worker. This was only an initial point, not foreseeing other forms of interaction except, indeed, to avoid danger in safeguarding humans. Modern HMI, based on AI, now merges with the concepts of HCI (Human Computer Interaction), the discipline oriented towards increasingly natural forms of interaction with the computer, now increasingly endowed with its own form of intelligence. The transformation from the mouse to natural touch interfaces to upcoming gesture-based interactions is just one example. Natural interaction is now largely provided by language, which has enabled computers, machinery, robots, and cars to understand human language, respond, and also apply non-verbal forms of interaction (e.g., predicting actions, providing explanations on the context through recognizing the posture of workers involved in strenuous tasks, or facial expression) and sophisticated artificial vision systems capable of recognizing and segmenting the scene.

These new forms of interaction are already transforming daily work, allowing for greater freedom for individuals. For example, the ability to interface through speech – a natural mode for us – instead of writing or using a mouse, relieves the operator from tiring positions, always sitting and close to the keyboard or mouse, and also allows for parallel interactions with other tasks (e.g., while moving). Increasingly intelligent interactions will also be enabled by the combination of multiple sensors (microphones, GPS, proximity sensors), allowing for the geolocation of the operator. These forms of interaction are changing office work as well as industry and naturally require reskilling for workers, but they also go in the direction of making intellectual and manual work less fatiguing.

Artificial Intelligence is a tool for cooperation with human work.

The concept of cooperation goes beyond mere interaction and presupposes the common intention to perform an action or develop a process in a joint artificial-natural manner. A clear example is the cooperation in new forms of learning of neural networks. In their first generations, of supervised networks, the human expert provided both the data and the expected output in an offline manner. Large datasets of input-output pairs were given to the learning algorithm. This form of cooperation – necessary especially in cases of specific know-how – is very labor-intensive, requiring the commitment of specialized labor, or domain experts, including the managerial level.

Now the form of cooperation has reached much higher levels. On one hand, AI systems can cooperate to provide data (e.g., synthetic data) useful for better learning, so much so that a widely used paradigm now is "teacher-student" or "knowledge distillation," where a neural network learns not only from data provided by the operator but also from an already trained network that provides knowledge. A similar concept is found in "Domain adaptation," where a general, pre-trained network is used as a starting point along with other information provided by the expert to adapt to a specific domain. For example, a network capable of understanding anomalies in time series can be adapted to a domain of energy production, or traffic anomaly reviews, with cooperative learning. Conversely, the already mentioned form of RLHF – largely unconscious – has the direction from person to machine, used by new LLMs to continuously improve their effectiveness by collecting user responses and reusing them for new learning iterations. Training and awareness of workers will be increasingly necessary to make these forms of co-learning effective, but also correct and ethically consistent.

But even more important is the form of cooperation at the level of inference with already trained systems, especially in generative AI, which constitutes the real pivot of the AI-based Economy.

The pillars of the Knowledge-based economy at the beginning of the 21st century (despite the definition deriving from economic texts of the 60s) were specific knowledge, human capital, and intangible assets such as intellectual property, the ability to propose non-standard actions and solutions on the market, due to the specialized knowledge of sector experts, supported by technology, also oriented towards customization, like 3D printers, and digitalization. The OECD has particularly pushed as a think-tank to spread these paradigms towards the development of high-tech companies strongly based on human capital [Godin, 2006)] spread. These pillars have remained solid even in the years to come but in synergy with new types of economic actions based on the analysis- also automatic- of Big Data. The Data-driven economy has been much theorized in the past decade as an economic system in which the production of goods and services is mainly based on intensively data-driven activities. In fact, for several years economic production has been based on the use of large amounts of data, structured, unstructured, interoperable, and in any case interpreted by managers and technicians capable of using sophisticated business analytics tools also for decision-making. This is the digital evolution of the context in which business decisions were mainly based on experience and knowledge accumulated by management. [Cavanillas et al 2016]. This period of diffusion of the big data paradigm - however analyzed by human experts, always with computer tools- as a driver of production and the evolution of decision-making processes, has now given way to a new form of economy, which J. Maggioncalda, CEO of Coursera, the largest online learning platform, called the post-knowledge economy at the World Economic Forum in 2023 [Vitanova et al 2023], indicating in the dynamic skills of operators, the exchange value in the new economy centered on AI. This need for reskilling and upskilling may affect 44% of workers in the next 5 years given the dynamism of the evolution of technologies and in particular of generative AI, which was the most striking example of rapid adoption in a very short time. The Chat GPT phenomenon, which in the space of one year from its appearance on the market has gained more than 180 million users and which in October 2023 collected about 1.8 billion monthly visits according to official statistics, is just the tip of the iceberg.

Although in-depth economic analyses are still premature, artificial intelligence as a new pervasive engine of the economy is already a topic of interest all over the world and also in Italy, both at the level of business associations, banking foundations, and public institutions: a neologism like 'AI-based economy' can help understand how we are moving in this new decade of the century towards an economic system in which the production and consumption of goods and services is increasingly based on AI, and on the form of implicit and explicit knowledge derived from such systems, especially currently of the generative type, capable of providing new models of representations of reality and cooperating with experts, technicians, and managers for the achievement of even complex objectives. Quoting Maggioncalda, 'Whether you're ready or not, almost anything digital is going to change and become AI-enabled.

Through tools like conversational LLMs, generative systems of texts and images, AI systems capable of proposing decisions and predictions, work has changed and will change completely in the near future. Just as the digital tool has been necessary until now to explicitly retrieve data and information, interface with machinery, analyze data, from now on this will be done indissolubly in a cooperative way between expert people and artificial systems. This will also constitute a great corporate asset because, being able to cooperate simultaneously with multiple workers, it will allow full and parallel management of corporate know-how, which will be increasingly stored in the memory and intellectual capacity of AI. How much of the decision-making, managerial, operational capacity of individuals will be replaced by the machine will again depend on work transformation through continuous training, so that AI is a tool of the person and not vice versa.

5. AI Supporting Various Productive Domains

The National Strategic Program for Artificial Intelligence 2022-2024, promoted by the Draghi government [Pres. Consiglio dei Ministri, 2021], which, although in the revision phase, has been the reference document so far for PNRR funding policies on AI, defined 11 thematic priorities to concentrate AI interests in Italy (Fig. 4). These 11 themes show how AI is truly ubiquitous and will be integrated into working life in all contexts.

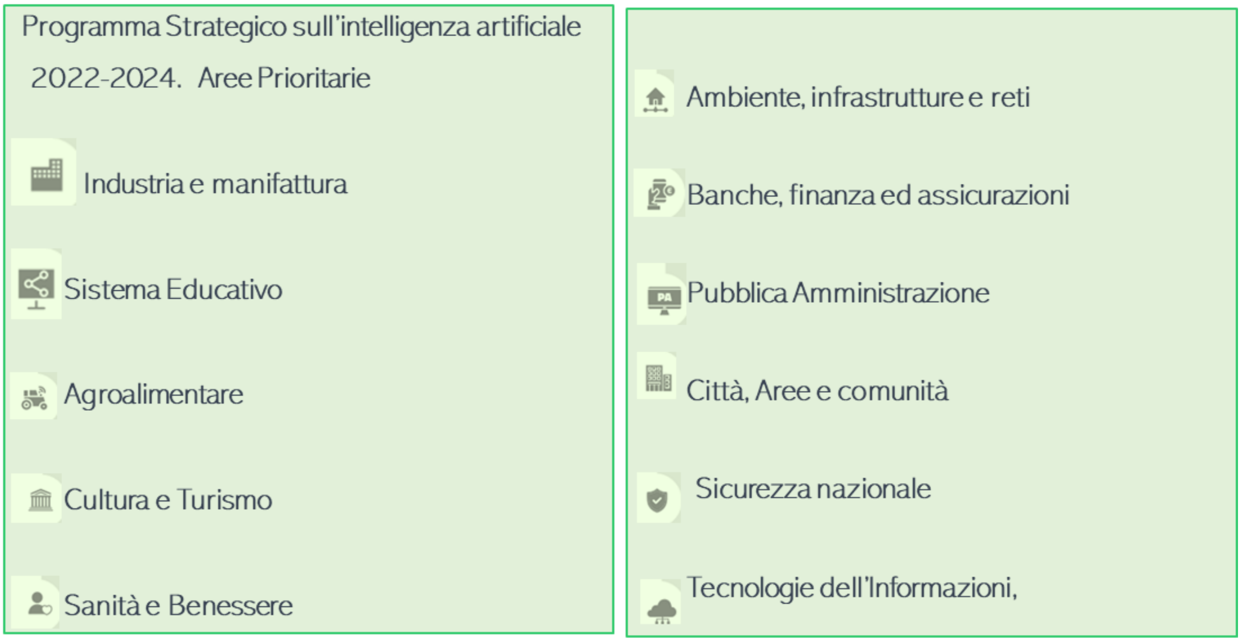

Fig.4 ***11 Priority areas in Italian Strategy ofn AI 2021-2024***

Primarily, AI is influencing the *world of industry and manufacturing*. This is a sector less highlighted by media spotlights and perhaps of lesser interest in the United States, but it is crucial for the economy of our country and Europe. AI can contribute significantly to the generational leap in production optimization, the design of new products (through generative design), energy saving, and maintenance with AI tools for predictive analytics. In this field, AI has not yet reached full maturity, still requiring the production of specific Machine Learning systems with specific data, and only large industries can afford extensive experimentation.

Similarly, there's a huge and growing impact in *the agri-food sector*, especially in precision agriculture, allowing for more sustainable solutions in terms of energy and chemicals. This area is expanding rapidly and will benefit significantly from improved collaboration between domain experts and computer scientists, along with related investments.

The use of AI in *the Education, Health and Wellness, and Culture and Tourism* sectors (so important for Italy) could fill pages of discussion, which are omitted here as they partly fall outside the scope of this document. Certainly, large-scale generative and conversational models are already leading to new solutions and services in these areas, especially for document analysis.

The macro-category of *Environment, Infrastructure, and Networks* is crucial. The cited Italian Strategy document emphasizes "the significant impact on resource conservation, emission reduction, traffic flow management and related risks, strengthening of the circular economy, and prevention of natural disasters." Additionally, there's the acceleration of ecological transition, the enabling role in 5G networks, and the management and monitoring of infrastructure, waste cycles, and hydrogeological instability control. This is an area still to be fully developed and one that could greatly benefit from the latest developments in Generative AI, both for document analysis and correlation, and for predictive analysis of time series for energy forecasts, anomaly analysis, or load distribution optimization. Experimentations are already underway in many large companies, such as the ENI-IBM collaboration since 2019. A specific concern is the energy consumption impact of AI, as highlighted in [OECD AI Footprint, 2022] and other documents. The development of AI for automatic training is very energy-intensive, requiring joint research and computing infrastructure efforts for future generations of sustainable AI. This topic is addressed by the European Project ELIAS [ELIAS, 2023].

The *Banking, Finance, and Insurance* sector, enriched by AI, is now one of the main areas of deployment and development of intelligent services, both for discriminative models for classification and risk analysis, and for generative models (including synthetic data generation to work without violating privacy aspects), and for the use of LLMs in customer interfaces and service personalization. American companies like JPMorgan and Salesforce are pioneers in the use of AI in finance, but in Italy too, companies such as Prometeia or CentAI are dynamically and effectively employing generative AI.

National Strategy topics for AI applications in *Public Administration, Smart Cities, and Communities* are numerous and will be further developed in the future as reference themes for the transformation of work and efficiency results. AGID has recently declared its intention to act quickly to implement policies and programs for the massive use of AI in Public Administration, as in other countries. Among the many applications in the field of smart cities, in addition to mobility and traffic, the construction sector is noteworthy, where recent Machine Learning tools will be indispensable for the design and maintenance of large infrastructures, bridges, roads, airports, including for simulation purposes.

Emphasis should also be given to issues concerning *National Security*. Unfortunately, AI is a key tool both for undermining national security (as a perfect tool for cybercrime) and for combating its effects. Efforts are also being made to combat other cyber threats, such as AI-generated fake detection (a specific topic of the European ELSA project coordinated by CISPA in Germany, with several Italian institutes participating, including UNIMORE and Leonardo), and physical security, surveillance, and biometrics, areas where AI has achieved striking results but also requires more regulation.Finally, a sensitive topic is the use of AI in DEFENSE and as a tool for LAWS (Lethal Automatic Weapon Systems), which is debated due to the difficult international situation, both technologically and in terms of regulation.

In Italy, the Information Technology sector is perhaps less developed compared to the American and Asian worlds. It's crucial for Italy to recognize this priority to demonstrate that it's not feasible to leave the domain of national AI research and production solely to foreign companies. The Italian situation is not ideal; many software companies and startups are growing in the country, particularly in Emilia-Romagna, though they are often acquired by foreign partnerships. This sector will need not only capital investments but perhaps also support and protective measures to ensure that human capital remains in the country, despite lower salaries and relatively limited opportunities compared to neighboring countries.

In all these domains, AI is already being experimented with and adopted, especially the new Generative AI. According to McKinsey [McKinsey, 2023], 60% of organizations worldwide are using generative AI in their work processes. In terms of LLM and generative systems, Europe is quite aligned with the rest of the world: 83% of organizations in China have tested it (19% use it regularly); in North America, 78% of organizations have considered it in testing (28% use it regularly); in Europe, 79% of organizations have tested it at least once (24% of companies use it regularly). These are significant results considering these tools have been known or in production for less than two years. Furthermore, the study reports that more than two-thirds of respondents expect their organization to increase AI investments in the next three years.

6. AI in Support of Intellectual, Managerial, and Organizational Work

Given the priority areas, which truly demonstrate the ubiquity of AI tools, it's easy to see how many of the highlighted tools are common to all productive areas and relate to the impact of AI as a tool in the world of work, in every intellectual activity in the organization, in management, and in various service stages.

In all these areas, there is a fear of the substitution of human activities due to the high degree of automation possible. Certainly, highly specialized roles, both with AI skills and in the economic and managerial field, will not be affected, and simultaneously, lower-skilled roles can increase their productivity and job satisfaction with the cooperation of AI tools. According to various economic studies, the most at-risk group is the middle tier, where skills will need to be more specialized to avoid being automated by different AI tools, including generative ones. In the previously mentioned OECD report, interviewing different worker categories, about 27% of jobs across the OECD average are feared to be automated (with high variance: less than 20% in countries like Luxembourg, the UK, and Sweden, and more than 35% in countries like Slovakia, Hungary, and the Czech Republic). The report garnered interest in Italy because the value there is above the average (around 30%). The reasons, in my opinion, are multiple, but primarily due to a certain lack of information in our country, the higher average age of Italian workers, and a real lack of adequate skills, considering that our country is unfortunately at the bottom of the OECD in terms of the number of graduates.

Beyond concerns, AI in the service world is spreading and will necessarily be not a tool of substitution or pure aid, but a tool of cooperation. Powerful new tools like Language models require interaction with domain experts (for the part defined as "prompt engineering," i.e., the construction of inputs for conversational systems) to achieve acceptable results, both in terms of time and sustainable consumption. Similarly, in many contexts- such as finance- only the development of neuro-symbolic systems or the combination of purely data-driven systems with rules and knowledge obtained from human experience yield the most reliable results. In any case, AI as a work tool is cooperating with people in many areas such as the following (partially drawn from [Newmetric 2023]):

a. *Work Automation:* AI can be used in "workflow optimization," with functions now addable in CRMs and scheduling services to learn how to optimize business organization; the use of "predictive models" to foresee potential risks and bottlenecks in work and to streamline flows (like warehouse supply) is not yet widespread in Italy but is already internationally.
b. *Personalization*: Increasing tools are available to provide employees with "personalized learning" to identify skill gaps and customize the needs of different employees, making their growth more satisfying, and with "intelligent assistants" for which assistance tools are defined.
c. *Decision-making Enhancement*: In this area, cooperation is higher because AI tools, both discriminative and generative, can now assist in "real-time data retrieval" as an analysis, summarization, and

document correlation tool, and in "real-time decision support" providing similarities with the past and suggestions for human decision-making.

d. *Collaboration Improvement:* Tools already exist to provide "real-time feedback" in conversations, online meeting results, and "collaborative rooms and resources" to bring the best work tools together. This includes co-pilot tools for software code generation (an area where tools like ChatGPT and others provide the most reliable results when used by workers with appropriate skills).

e. *Work-Life Balance Improvement*: Tools are available to assist in tasks such as "prioritization" of activities, "time management" to schedule the best meeting times, and to optimize "remote work" by providing better collaborative tools.

Many of these tools are in the experimental phase but are recently settling and will provide productivity improvement tools. A downside, always highlighted in the OECD document and many recent analyses, is that these tools are cognitively demanding and can lead to workload fatigue if not known and used properly.,

7. Striking a Balance Between Risks and Opportunities

I tools are revolutionizing work, individual lives, and society, and create considerable apprehensions due to their potential power, concentration in the hands of a few manufacturers (Google, Amazon, OpenAI, Salesforce, IBM, Apple, Microsoft to name a few in the Western world, and corresponding companies in the Asian area like Tencent, Baidu, etc.), and also for their lack of transparency.

Transparency, as defined in previous sections, is indeed a problematic aspect: a) the need to recognize whether data are generated by machine or human; b) the need for interpretability of results and the motivations leading to decisions; c) the need for knowledge of the learning and inference process (for example, knowing the data used for learning). A recent extensive document from Stanford's Human AI Center showed that the transparency level of the 10 most famous current generative systems from the aforementioned companies is still extremely lacking in the area of transparency [R. Bommasani 2023].

In every domain, to hope for effective AI diffusion in both its adoption and development, a revision of every AI system, classic, vertical, or foundational, in a "Trustworthy" key, as defined in previous pages, will be necessary. If the opportunities are clear, so are the risks, and they must never be underestimated. In particular, the capabilities of LLMs, even in open-source forms, will soon be put to the test in important moments for international democracy, such as the upcoming European and American Elections.

Demands for Responsibility and Reliability by design and by default of AI systems are being affirmed at the international governmental level. On one hand, they are one of the foundational elements of the already mentioned AI Act, a document that passed the first level of approval in the European Community in the first half of 2023 but requires a lengthy parliamentary process, both at the European level and then of the member states, to be accepted and which, to date, has not found convergence on every issue, especially on the responsibility of foundational models. The AI Act does not discuss the technology but the risks of applications, defining "unacceptably risky" applications, which should be banned in the European community (especially concerning the unconscious profiling of individuals), others at high risk that will have to obtain a CE mark (as in the educational or health sector) and precise certification for risk management, and others that, while of contained risk, must still maintain the tenets of transparency.

The dichotomy between maximum freedom for optimization of opportunities and maximum regulation for risk mitigation is the prevailing theme of recent times. Unfortunately, technologies are international and, as mentioned, dominated by international and private actors whose transparency is not guaranteed. One example among many is the recent events with OpenAI CEO Sam Altman, who was removed from his position, hired by the largest funding company, Microsoft, and then reinstated with a new board, within a few weeks, for reasons related to the transparency of activities in the company.

The necessity for regulation is acutely felt on a global level, and for the first time, a global meeting was held at Bletchley Park, resulting in the Bletchley Park Declaration signed by 28 states, including Italy. This declaration was the first official response to the need for clarity and transparency for safe AI, not only for the widespread use by individuals in everyday life and at work but also for the development of future AI.

The declaration is a milestone, but even more so a starting point for the awareness that in the future, different digital and human intelligences will necessarily need to cooperate for a better and sustainable future of the planet and humanity itself.

8. Conclusions

This document was created as a point of discussion for greater awareness, even for non-experts, of AI's potential in work collaboration. Defining three conclusive points on which there can be no equivocation, it's essential to highlight: a) Education, b) Responsibility, c) Investments.

Education is a key point both for spreading among workers for the acceptance of systems and for non-prejudiced understanding of risks and benefits, but even more so, technical and scientific education is necessary for new generations and re-education for those already in the workforce who have only grazed the transformation of the AI-based Economy. Impactful actions like the recent ones in Finland [Shapiro 2022] at the country level, but also capillary actions at the level of individual companies and institutions, are necessary.

Responsibility can be interpreted in various directions: on one hand, AI and the systems themselves must be designed and implemented to be responsible, i.e., reliable and robust, transparent, and explainable, possibly free of training bias, increasingly frequent and sometimes uncontrollable in large-scale systems. Responsibility also stands for accountability, i.e., a legal responsibility of AI producers, which has been somewhat neglected until now, and the responsibility of institutions called upon for regulation.

Investments, finally, is an aspect not discussed so far but necessary to mention in the conclusions. AI, especially generative AI, implies huge investments: Microsoft invested over $40 million in 2023 in co-financing OpenAI, and Amazon recently invested $4 million to acquire a "startup" like Anthropic just to keep pace with other Big Tech companies. Investments in our country cannot be only in the purchase or use of platforms, or in financing computing resources, but also in investing in human capital (with European or international salaries) to make Italy a producer and not just a consumer of Artificial Intelligence.

Bibliography and Web references

[K.Shlawab, 2015] Klaus Schwab, The Fourth Industrial Revolution, World Economic Forum, 2015

[EU AI Act, 2021] European Parlament "AI Act" EU AI Act: first regulation on artificial intelligence https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[G7 Hiroshim,a 2023] "Statement pf the Hiroshima AI Process" 30 Oct 2023 https://digital-strategy.ec.europa.eu/it/library/g7-leaders-statement-hiroshima-ai-process

[Bletchely, 2023] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/

[IBM The FourV, 2012] IBM, The Four V's of Big Data, da http://ibmbigdatahub.com , Agosto 2012

[O Hersent et al., 2012] Olivier Hersent, David Boswarthick, Omar Elloumi. The Internet of Things: Key Applications and Protocols, Wiley; 2 edition 2012

[D.Ciurial, 2017] Dan Ciuriak The Knowledge-based and Data-driven Economy Quantifying the Impacts of Trade Agreements CIGI Papers No. 156, Dicembre 2017

[EU AI for Europe, 2018] EU Parlament "Artificial Intelligence for Europe COM/2018/237 final 2018 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

[OECD Legal 2023] OECD Recommendation of the Council on Artificial Intelligence 2019 with amendments in 2023 https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[R. Bommasani et al. 2021] Rishi Bommasani et al. "On the Opportunities and Risks of Foundation Models" Stanford report 2021 https://crfm.stanford.edu/report.html

[PWC. 2019] PWC Artificial Intelligence https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html , 2019

[Fortune, 2021] Artificial Intelligence and the market https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114

[McKisney, 2023] "The state of Ai Generative" https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai%20in%202023%20generative%20ais%20breakout%20year/the-state-of-ai-in-2023-generative-ais-breakout-year_vf.pdf

[Vaswani et al, 2017] Ashish Vaswani, et al. "Attention Is All You Need" Proceedings of Neurips 2017

[Stefanini et al, 2022] M Stefanini, M Lovino, R Cucchiara, E Ficarra Predicting gene and protein expression levels from DNA and protein sequences with Perceiver, Computer Methods and Programs in Biomedicine, 2023

[S.Luetto et al., 2023] S Luetto, F Garuti, E Sangineto, L Forni, R Cucchiara One Transformer for All Time Series: Representing and Training with Time-Dependent Heterogeneous Tabular Data ArXiv preprint arXiv:2302.06375 2023

[A.Palazzi et al. 2020]. A Palazzi, D Abati, S.Calderara F Solera, R Cucchiara Predicting the Driver's Focus of Attention: the DR (eye) VE Project, IEEE transactions on pattern analysis and machine intelligence 41 (7), 1720-173

[LR.Cucchiara et al. 2020] R Cucchiara, S Calderara, F Lanzi, A Mariani, E Pellesi, Generative system for the creation of digital images for printing on design surfaces US Patent App. 17/792,73

[Holzinger et al, 2022] A Holzinger, M Dehmer, F Emmert-Streib, R Cucchiara, I Augenstein, Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence, Information Fusion 79, 263-278 2022

[Godin, 2006] B.Godin, The Knowledge-Based Economy: Conceptual Framework or Buzzword?. J Technol Transfer 31, 17–30 2006

[Cavanillas et al 2023] J.M. Cavanillas, E. Curry , W. Wahlster Editors New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe Springer 2016 AI economy

[Vitanova et al 2023]B. Vitanova, K. Sampatar How to leverage your skills in the post-knowledge AI economy, Word economic formum 2023 https://www.weforum.org/agenda/2023/04/how-to-leverage-skills-in-post-knowledge-ai-economy/

[Pres.Cons Ministri, Strategia 2021], Programma strategico per l’Intelligenza Artificiale 2022-2024 https://assets.innovazione.gov.it/1637777289-programma-strategico-iaweb.pdf

[Iotanalytics 2023] https://iot-analytics.com/number-connected-iot-devices/

[Insiderintelligence, 2023] https://www.insiderintelligence.com/insights/drone-industry-analysis-market-trends-growth-forecasts/)

[G. Hinton, 2023]. Geoffrey Hinton The Future of Intelligence, intervista al MIR 2023, Youtube https://www.youtube.com/watch?v=sitHS6UDMJc

[Eu White Paper, 2020] Eu Commission, White Paper on AI, an Ecosystem of excellence and trustworthy 2020

[R.Cucchiara 2021] R. Cucchiara, “ L’intelligenza non è artificiale” Ed. Mondadori 2021

[OECD, 2023] OECD Employment Outlook 2023, Artificial Intelligence and the Labour Market

[OECD AI Footprint 2022] “Measuring The Environmental Impacts Of Artificial Intelligence Compute And Applications The Ai Footprint” n. 341 OECD Publishing 2022 https://www.oecd-ilibrary.org/docserver/7babf571-en.pdf?expires=1699209464&id=id&accname=guest&checksum=AC1B3CD9881F1A154798E4753C3D54A6

[ELIAS, 2023] European Lighthouse for AI Sustainability https://elias-ai.eu/

[Newmetric 2023] The future of work AI and its impact on employee experience 2023 https://www.newmetrics.net/insights/the-future-of-work-ai-and-its-impact-on-employee-experience/

[R. Bommasani 2023] Rishi Bommasani*1 Kevin Klyman*1 Shayne Longpre2 Sayash Kapoor3 Nestor Maslej1 Betty Xiong1 Daniel Zhang1 Percy Liang1 The Foundation Model Transparency Index Stanford HAI Publicaton 2023

[H. Shapiro 2022] H. Shapiro “Finland: AI, policy innovation and the future of work and learning” Denmark 2022 https://digitalfuturesofwork.com/wp-content/uploads/2022/10/Working-Paper-7-Finland-AI-policy-innovation-and-the-future-of-work-and-learning.pdf