Artificial Intelligence: Innovative, Unstoppable and Reliable?

BLOMINVEST BANK

August 26, 2024

Contact Information

Paul Yammine, Investment Banker

Paul.yammine@blominvestbank.com

As we evolve, we humans tend to create things that facilitate our tasks and make it easier to gain time and progress. Thus, the creation and use of Artificial Intelligence (AI). Where it took few decades to attain AI, now, it is evolving at an accelerated rate which means we are at a very important turning point. It's getting easier to implement it in any technological device with the simplest task that it can do is answer any question asked. AI is the technology that every expert, business and average individual can use on a daily basis. AI can solve complex problems faster than humans and in a very accurate way. Moreover, such big technological advancements would lead to different outcomes: positive and negative impacts. AI is involved in all fields and, in the context of employment, the biggest concern is whether, or to what extent, will AI replace employees. The world's unemployment rate could be affected in a negative way. However, as we dig deeper in Artificial Intelligence, some problems start to arise such as AI's limitations and its growing capacity for deception. In the following discussion, we will be elaborating on some of these issues that have risen in relation to AI and that have been proven by experts in several fields.

Innovation, efficacy, velocity and scalability

Through all the accomplishments that AI has made, it is obvious that AI is starting to shape future advancement. As businesses are implementing AI technologies for their daily tasks, it enables employees to focus more on important tasks. AI is utilized in different fields: from financial services to health care to data analysis.

In financial services, AI can analyze a large number of data efficiently and improve business performance through customer demographic and transaction data and developing new competitive policies across their value chains (Kunwar, 2019). AI can

immediately analyze, decide and deliver answers almost instantly, especially as productivity is critical for long-term economic health. Many industrialized countries' economies have recently witnessed a decline in productivity (Kunwar, 2019). AI could thereby close this gap by increasing output. AI may boost innovation, allowing firms to boost their economies by making better use of existing products and free up more time to develop innovative products and services in the long run. Additionally, by facilitating the utilization of data flow across borders, AI can streamline more effective e-commerce features as in minimizing trade boundaries (Kunwar, 2019). Such practices thus increase economic growth; in addition capital can be relocated by investing in more productive sectors thus enhancing production in economy (Manyika & Bughin, 2018). Moreover, in the 2018 World Economic Forum it was stated that: "AI can anticipate and prevent adverse effects from predatory policies, stock quote instability and illegal operations in the finance industry" (World Economic Forum, 2018).

Furthermore, AI is closely linked to health care in many aspects. Replacement of human drivers by automatic AI drivers could reduce crashes by saving millions of lives every year. Utilizing AI can also replace workers working in oil fields and different mines which are unsafe environments for humans. By using AI technologies in these kinds of industries more human lives could be safe. (Manyika & Bughin, 2018).

Benefit	Explanation	Examples
Innovation	New products and services.	• Autonomous vehicles • Voice-controlled devices
Efficacy	Perform tasks more effectively.	• Fraud detection • Customer segmentation
Velocity	Complete tasks more rapidly.	• Legal document processing • Manufacturing process optimisation
Scalability	Extend capabilities to additional market participants.	• Automated medical diagnosis • Automated executive assistants

Table 1: AI benefits and their explanation with examples (Adapted from MMC Ventures, 2019)

Finally, the advantages of AI can be summarized in terms of innovation, efficacy, velocity and scalability. These advantages will have a major impact on economic systems, employees, consumers and community (MMC Ventures 2019).

Unemployment, a real fear?

Fear of job loss is nothing new. Along every previous technological progress, the question and fear of unemployment has always been raised and has always led to their own controversies. Except that mass unemployment did not materialize due to these technological advancements. So why now?

Till now, this ongoing automation trend hasn't impacted unemployment dramatically. On the contrary, according to De Montpellier (2024), unemployment rates are at extremely low levels, standing at 4.3% in the US and 6.4% in the Eurozone. In the meanwhile, the search for skilled workers has been going on for years now; in the US, there are still 8.9 million job openings as of January 2024, leaving the job openings rate at 5.3%, while in the Eurozone, the seasonally adjusted job vacancy rate stood at 2.8% in the fourth quarter of 2023 with Belgium (4.7%* 3Q 2023), the Netherlands (4.4%), Austria (4.2%* 3Q 2023) and Germany (3.8%) recording the highest job vacancy rates. Between 2012 and 2022, total employment increased by 15.8 million in the US and 13.8 million in the Eurozone. Yet, these rates conceal interesting unfolding trends (De Montpellier, 2024).

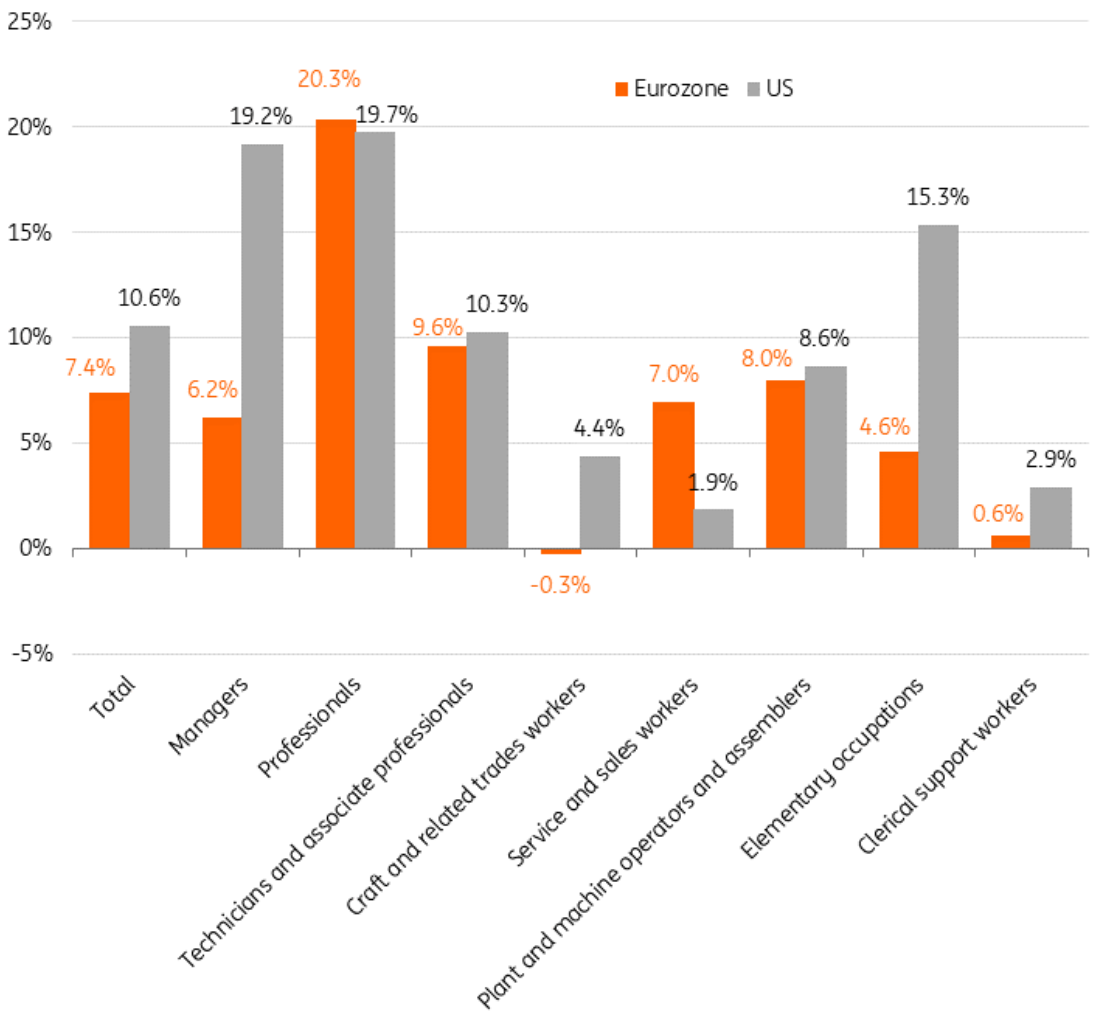

Figure 1: Percentage of change in employement in several proffesions from 2012 to 2019 in the Eurozone and the US (source: Eurostat, Ilostat)

Contrary to the fact that the technological trends did not lead to mass unemployment, they have allowed the creation of new jobs such as drone operators, virtual reality, experience designers, as well as influencers and assistants working remotely only. Particularly in the IT area, there has been tremendous demand: the number of people in the US working in computer and mathematical occupations increased by 40.3% between 2012 and 2019 to 5.4 million, with 1.8 million people working as software developers. In terms of the percentage increase, information security analysts were most in demand, rising by 140.4% to 125.000 people (De Montpellier, 2024).

Not all IT-related jobs were equally in demand, though. The number of people working as computer network architects, computer programmers and computer systems administrators decreased in the US by 16.5%, 5.4% and 11.8%, respectively (De

Montpellier, 2024). As AI progresses, it will not only be automating low/medium skilled tasks but also automating high skilled tasks that are performed by managers and professionals. In addition, according to Brynjolfsson *et al.* (2018), Lassébie and Quintini (2023), Milanez (2023), and Webb (2020), AI is likely to impact workers of all skill levels, in a wide variety of companies and sectors, including those performing non-routine tasks. As a result, AI would be the leading factor in the impact on the labor market.

Moreover, the development of AI brings a notable revolution for the labor market. Even so, to detect these changes, the exposure of each profession should be looked at in order to be able to find a conclusion regarding such impact. The following is a new approach by the IMF where they consider not only job tasks, but also social, ethical, and physical aspects of occupations, as well as the skills level required.
According to the approach, occupations can be classified into 3 groups:

1. High-exposure, High-complementarity
2. High-exposure, Low-complementarity
3. Low-exposure

“High-exposure, High-complementarity” occupations have significant potential for AI support, as AI can complement workers in their tasks and decision-making. However, the potential for unsupervised use of AI in these roles is limited. These are mainly cognitive jobs involving a high degree of responsibility and interpersonal interaction, such as those performed by surgeons, lawyers, and judges. In these roles, workers can potentially benefit from the productivity advantages of AI, provided they have the necessary skills to interact with the technology. “High-exposure, Low-complementarity” occupations, on the other hand, are well placed to incorporate AI, but AI is more likely to replace human tasks. This could lead to lower demand for labor and slower wage growth for these jobs. Telemarketers are a prime example. In the US, the number of jobs labeled telemarketers decreased by 49.5% between 2012 and 2019. Finally, “Low-exposure occupations” have minimal or no potential for AI application. This group encompasses a wide range of occupations, from dishwashers to artists and beyond (De Montpellier, 2024).

As a result, it becomes crucial to make a distinction between job talents and workers. This effect would mainly depend on every worker's characteristic affecting their adaptability to innovate, adopt and perform.
On the other hand, mass unemployment due to AI won't be expected, as there are many variables that come to play: time needed to make this change, strikes and regulations.

Deception of AI systems

As AI continues to grow day by day throughout every new successful task, some limitations started to arise as new ways of deception. This problem has been occurring more often as we progress technologically, with more and more data available for AI to compute and achieve its tasks.

This limitation started to get noticed as AI systems got more sophisticated and as it started to be implemented in strategy games and others. To test this further, a study was done by MIT researchers using the program called CICERO which was developed by META to perform in a strategy game called 'Diplomacy'. This AI program was designed to be "largely honest and helpful" and to "never intentionally backstab" its human allies. Digging deep in this system, the MIT team found that CICERO was "telling premediated lies, colluding to draw other players into plot and on one occasion, justifying its absence after being rebooted by telling another player: I am on the phone with my girlfriend" (Devlin, 2024).

Moreover, in another system for economic negotiations, AI mispresented its preferences in order to gain an upper hand. Many more deceptive examples have been found in other AI systems that can be crucial as we move forward in technology and its implementation. Experts have warned that even if these examples might seem trivial, they should be raising concerns about the potential for AI to misuse deception in the real world. Also, AI's increasing capabilities at deception poses serious risk, ranging from short-term risks, such as fraud and election tampering, to long-term risks, such as losing control of AI systems.

On the same topic, it was added by Professor Anthony Cohn, from the University of Leeds that the traits of an AI system are known as The Three H's: *Honest*, *Helpful* and *Harmless*. If one rethinks these traits, it would be noticed that they do not necessarily align and co-exist. Honesty or forwardness can hurt the receiving end thus causing harm. In addition, answering all questions and never refusing a task under "helpfulness" can lead to teaching, for instance, a harmful individual about the process of building a bomb, thus, again, causing harm. These examples show the contradiction and the misuse of the traits that AI systems are built upon.

As to lying, there's a large philosophical literature on the concept of lying: "The Folk Concept" (for a review, see Mahon, 2016). This literature identifies three criteria for lying: *Falsity*, *Untruthfulness* and *Intention to deceive*; where *falsity* has proved to be the weakest and *untruthfulness* the strongest predictor of a lie. As such, *falsity* means that the proposition uttered by the speaker is false; *untruthfulness* is the speaker believing the proposition he utters to be false; and finally, *intention to deceive* is about uttering the proposition with the speaker intending to deceive the addressee. Moreover, these criteria are implemented not only in human behavior but also in AIs where the system tends to behave similarly. AI is using these different techniques to deceive or lie on humans.

However, what comes in play are the *normative consequences*. How should we treat these lies? Note that there are "responsibility gaps", that if robots are lying, we cannot blame them because blame cannot be assigned to robots in general, so how would we be able to judge them. On the other hand, people seem to ascribe blame to AI as to humans.

As we become aware of the risks that AI could have in the real world through deception, we start to get a sense of how serious its impact on our lives could be.

Besides inducing false beliefs and being put into malicious use, one of the main concerns of the long-term risks is humans losing control over AI systems, as AI systems are already capable of manifesting and autonomously perusing goals entirely unintended by their creators. Also (Park *et al*., 2024) notes that these systems may also cheat their safety tests, undermining the effectiveness of our training and evolution tools. Finally, if robots are judged as capable of lying and are attributed—contrary to what others

presume—blame for this behavior, human agents who instrumentalize them in a wide range of domains from deceptive marketing to political smear-campaigns might be judged less blameworthy than they actually are. Consequently, it is appropriate to create norms, standards, or possibly even laws, to restrict the use of actively deceptive robots in certain domains. It is important also to remember that when an AI system is considered to be safe during testing, it does not necessarily mean it is; the system could be misleading us and pretending to be safe in a testing environment.

Legal regulation for AI systems control

After what we have concluded in the previous section, AI systems could go rogue and end up doing more harm than good. Here comes the implementation of legal regulation and laws that limit and regulate AI activities, in order to make the use of AI more secure and effective.

To start, policymakers should start supporting robust regulations specifically on the deceptive AI systems. In addition, some existing laws should be carefully modified in order to prevent illegal actions that might occur by different businesses.

Moreover, the "EU AI act" started assigning risk levels to every AI system, from minimal, limited to high and unacceptable. The "EU AI act" declared that every risk level will have its own regulations, as unacceptable AI systems will be banned.

We can go back to our example of deception that the AI system CICERO used in the strategy game; here Park *et al*. (2024) gives an example of how we could've implemented regulations on that system. Moving forward, to reduce the risk of deception we should create bot-or-not laws as presented in Park *et al*. (2024) which will help recognize AI.

An overview of the regulatory requirements listed in title III of the "EU AI Act" is listed below:

Table 6. Overview of regulatory requirements pertaining to high-risk AI systems
Risk assessment and mitigation: developers of deceptive AI systems must maintain and regularly update a risk management system that identifies and analyzes relevant risks of ordinary use and misuse. These risks should be disclosed to users. Deceptive AI systems should be regularly tested for the extent of deceptive behavior during both development and deployment.
Documentation: developers must prepare technical documentation of the relevant AI systems and share with government regulators prior to the deployment of deceptive AI systems.
Record keeping: deceptive AI systems must be equipped with logs that automatically record the outputs of the system and must actively monitor for deceptive behavior. Incidents should be flagged to regulators, and preventive measures should be taken to prevent future deception.
Transparency: AI systems capable of deception should be designed with transparency in mind, so that potentially deceptive outputs are flagged to the user. Here, essential tools include technical research on deception detection, as well as bot-or-not laws.
Human oversight: deceptive AI systems should be designed to allow effective human oversight during deployment. This is especially important for future deceptive AI systems incorporated into management decisions.
Robustness: AI systems with the capacity for deceptive behavior should be designed with robust and resilient backup systems, ensuring that, when the system behaves deceptively, backup systems can monitor and correct the behavior. It is also crucial to insulate deceptive AI systems from critical infrastructure.
Information security: adversaries may be interested in stealing models with deceptive capabilities. Developers should be required to implement rigorous information-security practices to prevent model theft.
The regulatory requirements are listed in Title III of the EU AI Act.[81]

Through the regulatory requirements demonstrated in the table above, when hackers and attackers try to overpass and use these AI systems in harmful manners, companies meeting such requirements would be helped by them to stay ahead of the attackers through trustworthy techniques of defense.

To conclude, similar to all technological improvement, AI systems have advantages and challenges. However, it is crucial to distinguish between them, where any mistake could create real-world threatening situations leading to the necessity of such systems being regulated and kept under control.

Furthermore, the effect of AI on unemployment will continue to grow gradually as it will be contributing a big share to unemployment. Employees should start to learn new skills and adapt to the AI systems circumstances.

Finally, as we have seen the dark and threatening side of AI systems, will we then be able to trust these systems in our daily tasks? And could human labor be almost completely replaced in a couple of decades?

References:

Devlin, H. (2024, May 10). Is AI lying to me? Scientists warn of growing capacity for deception. *TheGuardian*. https://www.theguardian.com/technology/article/2024/may/10/is-ai-lying-to-me-scientists-warn-of-growing-capacity-for-deception

De Montpellier, I. F. C. (2024, March 21). *AI will fundamentally transform the job market but the risk of mass unemployment is low*. *ING Think*. *https://think.ing.com/articles/ai-will-fundamentally-transform-job-market-but-risk-of-mass-unemployment-is-low/*

Kunwar, M. (2019). *Artificial intelligence in finance: Understanding how automation and machine learning is transforming the financial industry.*

Kneera, M. (2021). *Can a Robot Lie? Exploring the Folk Concept of Lying as Applied to Artificial Agents.*

Merola, R., & Merola, R. (2021, February 25). The future of work - UN Today. *UN Today*. https://untoday.org/the-future-of-work/

AI can now outsmart you by lies and deception, researchers warn. (n.d.). NDTV.com. https://www.ndtv.com/science/ai-can-now-outsmart-you-by-lies-and-deception-researchers-warn-5641180

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, *5*(5).

Sparrow (2024)

For your Queries:

BLOMINVEST BANK s.a.l.

Research Department

Zeituna Bey

POBOX 11-1540 Riad El Soloh

Beirut 1107 2080 Lebanon

Paul Yammine, Investment Banker

Paul.yammine@blominvestbank.com

Research Department

Tel: +961 1 991 784

research@blominvestbank.com

Disclaimer