Department of Artificial Intelligence & Digital Marketing

Digital Agents and Business Innovation: How AI is Redefining the Rules

SUPERVISOR

Prof. Maximo Ibarra

COSUPERVISOR

Prof. Marco Francesco Mazzù

CANDIDATE

Carlo Policella (784901)

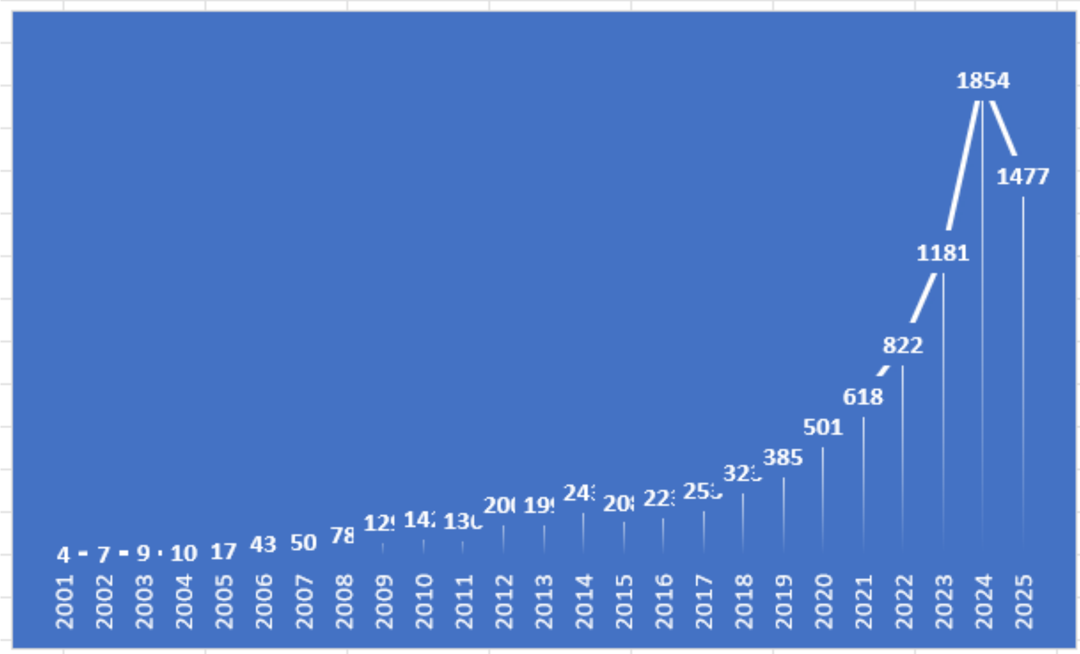
Academic Year 2024/2025

Summary

INTRODUCTION

Gartner, one of the leading strategic consulting firms in the information technology sector with over 15,000 clients worldwide, has identified Agentic AI as one of the 10 technology areas with the greatest potential for development in the near future in its *Strategic Technology Trends for 2025* report, within the AI Imperatives and Risks section. This is certainly not the first time that artificial intelligence has been included in Gartner's ranking. In the past, however, its inclusion among the most promising trends was less explicit: in the 2020 report, trend no. 4 was dedicated to Human Augmentation, understood as "[...] the use of technology to enhance a person's cognitive and physical experiences"; the report indicated AI as a technology capable of increasing humans' ability to think and make better decisions and highlighted its potential for development. In 2022, Generative AI was one of the topics identified as particularly promising for businesses. Gartner's 2025 Annual Report explicitly includes Agentic AI among the strategic technologies of the future, recognizing its importance and highlighting the continued growth of interest in this subject, which is no longer confined to academia or applied research. Its inclusion in a report that has been a fixture for Information Technology professionals for years was not an isolated event: in recent years, the number of scientific articles, research contributions, and operational applications related to AI agents has increased significantly, signaling an interest in applying the principles of artificial intelligence to many areas of human activity, including marketing. A search of the Scopus database based on the search string "Agentic AI" or "AI algorithms" or "Recommendation agents" returned 9,115 results from 2000 to the present; the number of contributions appears to be growing steadily over time, with growth becoming more pronounced since 2020 (Fig. 1).

Figure 1 – Distribution of scientific issues related to the topic of AI agents



Source: personal draft based on *Scopus*

This paper aims to explore the profound and rapid transformation that Artificial Intelligence (AI) and, in particular, its evolution into Digital Agents, are bringing to the world of business and marketing. Technological progress, once perceived as a linear and predictable evolution, has taken on an exponential dynamic in recent years, redefining not only operational processes but also strategic models and human interactions within organizations. From the birth of digital marketing in the 1980s, characterized by the integration of technologies such as floppy disks for commercial promotion, to the advent of big data and AI, the sector has shown an extraordinary capacity for adaptation and innovation. The consumer, once a passive subject, has transformed into an active "prosumer," a protagonist in a production process in which the distance between production and consumption is reduced, and personalization becomes a central element.

The main objective of this study is to analyze how the evolution of AI, from machine learning to generative AI and AI agents, is redefining the boundaries of what is possible in the business world. The focus is twofold: on the one hand, it aims to provide a detailed mapping of the architectures and functioning of AI Agents and , highlighting their key components such as the planner, tools, memory, and feedback loop; on the other hand, it aims to assess the practical, strategic, and social implications that their adoption entails for organizations. The work is not limited to describing the opportunities, but also focuses on the ethical, security, and governance challenges, offering a critical reflection on the evolving relationship between technology and the human factor.

The thesis is divided into five chapters, each of which addresses a specific aspect of this revolution.

The first chapter, (*Artificial intelligence in marketing and business*), presents a historical and conceptual overview of the topic. The starting point for the analysis is the definition of digital marketing, a prerequisite for correctly framing the stages of its evolution and understanding how the introduction of AI has been a profound innovation. The chapter highlights the ways in which AI has transformed marketing processes, enabling large-scale personalization and automation of operational flows. The impact of AI is clearly evident when comparing the substantial differences with traditional marketing in terms of accessibility, process monitoring, scalability, and conversion rates. The chapter illustrates the circular process that sees data analysis and AI as central elements in improving the offer of products and services: central elements that influence the interaction between the brand and the consumer, promising greater customer loyalty thanks to the ability of AI agents to influence their behavior and decisions.

The second chapter, (*From traditional AI to Generative AI and AI Agents*), traces the evolutionary path of Artificial Intelligence. Artificial Intelligence applications are linked to the development of Machine Learning and Deep Learning: these two approaches are given ample space in the chapter, as they are the theoretical and practical disciplines that have led to the revolution of Generative AI, which is described with particular attention to linguistic and multimodal models (such as Text-to-Text, Text-to-Image, Text-to-Video, etc.). etc.). Once the general framework has been established, a description of the characteristics of AI agents has been developed through two fundamental steps that emerge from the literature: first, a series of definitions have been presented, through which we have attempted to reduce some of the ambiguities that emerge from the literature; Finally, we structured a taxonomy of AI Agents, seeking to identify the different types based on their fundamental characteristics, thus laying the foundations for understanding their autonomous and interactive capabilities.

The third chapter (*AI Agents: architecture, operation, and business integration*) delves into the more technical aspects of the work, describing the structure and components that enable AI agents to function. The first part of the chapter analyzes the main modular components: the *planner* (the agent's brain), which formulates strategies and breaks down tasks; the *tools* that enable interaction with the external environment; the *memory* that stores short- and long-term information; and finally*,* the *feedback loop,* i.e., the process that enables self-assessment and continuous improvement of the AI agent through the analysis of responses from the surrounding environment. Based on these components, it has been possible to distinguish three types of AI agents (autonomous, collaborative, and hybrid). The chapter addresses how AI Agents enter the main production processes, interacting with business applications (CRM, ERP, and other systems that determine data flows), integrating data from a variety of different sources, and enabling decision-making processes. The last paragraph is dedicated to ethical, security, and governance implications: fundamental issues that define some of the most important challenges for the future development of AI Agents.

The in-depth analysis of ethical and legal implications forms the basis for the fourth chapter, (*Social and Organizational implications of the use of AI Agents*), which broadens the perspective from the purely technical and operational dynamics of AI agents to the social and organizational implications of their increasingly pervasive use. The impact of digital agents on the labor market is examined, analyzing both the positive and potentially negative aspects. The spread of AI agents appears to be characterized by a marked polarization between the phenomenon *of Enjoying Prosperity,* supported by those who emphasize with great optimism the advantages deriving from the use of AI not only in the rationalization of production processes, but also in the "quality of work," and the social phenomenon linked to the fear of many workers in production sectors most exposed to the use of artificial intelligence of being replaced (the phenomenon *of 'replacement anxiety')*. The chapter also investigates how the introduction of AI agents also implies a profound transformation of leadership within companies: classic models of directive leadership, fundamentally based on control, now seem inadequate to manage the change taking place; Instead, it appears more effective to develop a leadership style that facilitates human-machine interaction and is capable of addressing the broader social implications of the widespread adoption of these technologies.

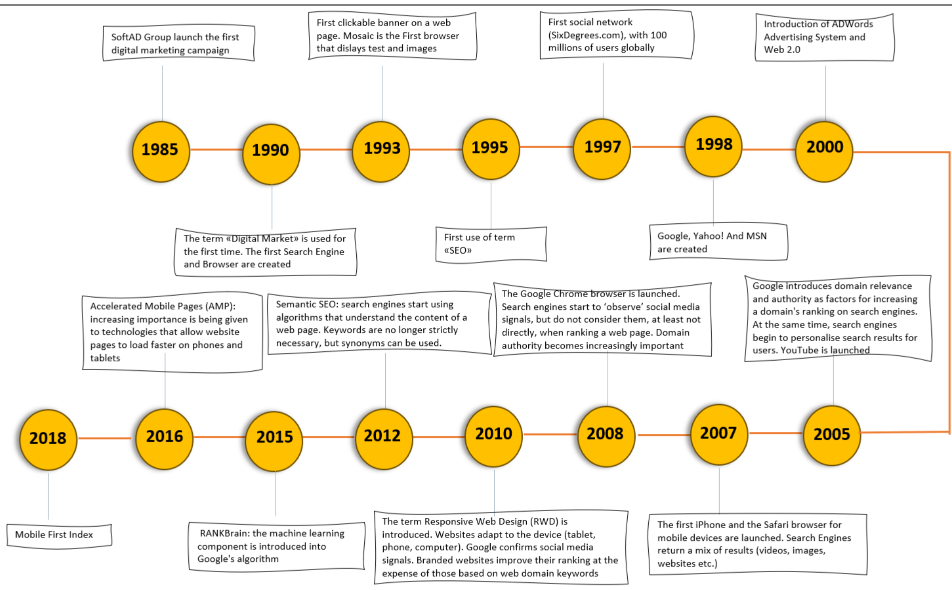
Finally, in the fifth chapter (*Comparative case studies*), we proposed some concrete examples related to the application of AI Agents in different business contexts. Several case studies were analyzed: Klarna's customer service management, promoted (not without criticism and second thoughts) through the replacement of human operators with AI Agents; the potential linked to the management of projects developed with AutoGPT; the optimization of the logistics chain and user experience (UX) adopted by IKEA; the integration of AI agents and, in general, artificial intelligence tools within business systems and, in particular, in CRM, analyzing the example and strategy of Salesforce Einstein. The comparative analysis of the different cases allowed us to evaluate "in the field" the elements we proposed in the theoretical analysis, verifying not only the efficiency and impact of AI agents on business performance, but also the ethical, commercial, and relational implications associated with the use of artificial intelligence.

1. ARTIFICIAL INTELLIGENCE IN MARKETING AND BUSINESS

1.1 Definition of digital marketing

The history of digital marketing has its origins in the 1980s, when the SoftAd Group launched advertising campaigns for automotive manufacturers based on the integration of traditional marketing and technology. Readers of specialised magazines could send a coupon cut out of magazines and in return they received a floppy disk with digital content and an invitation to a test-drive. It was an innovative idea that opened up the pioneering era of digital marketing, which would soon evolve from the 1990s onwards, when the term began to spread among specialists and began experimenting with the first operational application and increasingly sophisticated forms of communication: in 1993 the first clickable banner ad appeared and the following year the first web crowler capable of exploring, analysing and cataloguing web pages was realised (Bîrzu, 2023). The growth of digital marketing is linked to the spread and evolution of the Internet and the technological evolution and, since 2010, the massive use of mobile devices: in 2018, the speed with which web pages are opened on mobile phones and tablets has become the standard for defining the ranking of websites (*Mobile First Index*) (Fig. 2).

Fig. 2 - Key milestones in the evolution of digital marketing



Source: Personal Draft

Crucial to the digital evolution has been the exponential growth of network users from 16 million in 1995 (0.4 per cent of the world's population) to 5.65 billion in 2024, corresponding to approximately 68 per cent of the world's population (Digital 2025).
These brief notes on the evolution of digital marketing show how it is not simply a technological update of traditional marketing practices, but a profound redefinition of the marketing landscape, exploiting the pervasive reach and interactive capabilities of the Internet and associated digital technologies. The definition of digital marketing is affected by its link with technological evolution: also referred to as 'online marketing', 'web marketing' or Internet marketing', it encompasses a vast conceptual domain in which

> "[...] the use of technology, particularly AI, to automate and optimise marketing processes. It involves the use of algorithms, machine learning and other technologies to analyse data and make decisions about marketing campaigns' (Lakshmipriyanka et al, 2023).

It is therefore a multifaceted approach that encompasses search engine optimisation, social media marketing, content marketing, email marketing and pay-per-click advertising: due to the vastness of these areas, some literature has considered the concept of digital marketig an 'umbrella definition' that defines

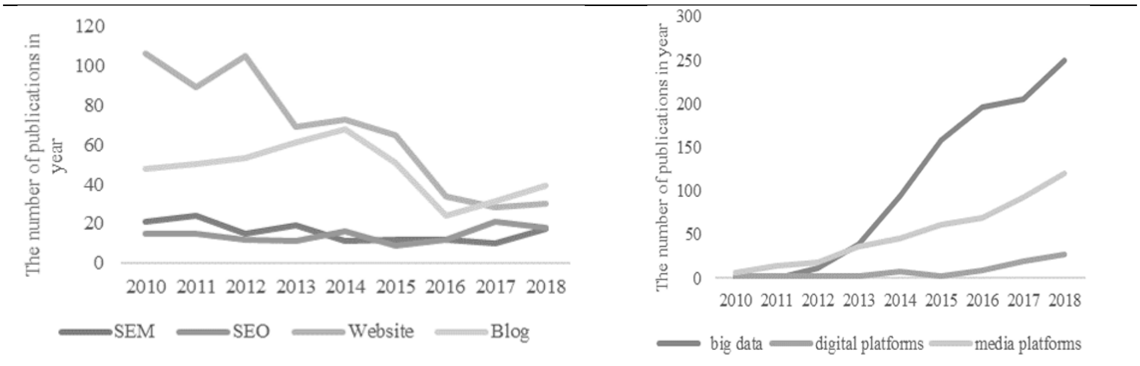
> "[...] the targeted, measurable, and interactive marketing of products or services using digital technologies to connect with current and prospective customers" (Choudahari et al., 2019)

The proposed definitions are very broad, but they show some common elements: the use of all the technologies available to marketers is basically aimed at extending the services offered and creating a meaningful connection between brands and customers: in an increasingly complex and competitive context, digital marketing strategies prove to be more effective because they go beyond the traditional linear approach in commercial communication to enhance the reciprocal, direct and open exchange with customers. Today, the consumer is no longer a passive object, but an active player in the production process in which the distance between production and consumption appears to be increasingly reduced. Today, the customer is increasingly a *prosumer*, a term introduced by Alvin Toffler in the 1980s: the term (which derives from the inlay of the terms *pro-*

ducer and *con-sumer*) highlights the intuition behind the current trend in modern marketing, which shows how companies are focused on the personalisation of the product or service (Toffler, 1987).

The shift to digital marketing reflects a broader trend towards digitisation in all sectors of the economy, driven by the increasing adoption of Internet-connected devices and the growing importance of online channels for communication, information access and commerce. This dynamic is also evident in terms of scholarly production: over the past two decades, researchers have shifted their focus away from traditional marketing topics and towards digital marketing, with the aim of identifying what future developments in the field might look like. A 2019 study based on an analysis of the specialist literature showed that articles dedicated to websites and blogs with marketing purposes are decreasing, while those related to SEO and SEM are essentially stable; on the other hand, there is a significant rise in studies based on big data and digital and multimedia platforms (Figure 3).

Fig.3 - Publication activity in different topics within the aim of marketing



Source: Alekseeva et al., 2019

The study just cited (Alekseeva et al., 2019) identified early on the tendency of marketing to expand its reach towards the use of platforms and especially on the use of big data. Indeed, one of the characteristics of digital marketing is the need for sufficient data to produce consistent processing. Data is essential for marketing because it allows marketers to tailor their strategies and tactics to specific audience segments with unprecedented precision. In a data-dominated society, those with the ability to aggregate, manage and utilise large amounts of data have an objective competitive advantage over others. The amount of data available on the network is enormous and growing: their collection, storage and processing takes place within what the literature, borrowing the language of the biological sciences, calls 'data ecosystems'. This term refers to the multiplicity of networks of organisations and individuals that feed, exchange and use data as a main resource, determining new business models. A data ecosystem is based on collaboration between the various stakeholders and their coordination with regard to the creation, management and sharing of data. Participation in a data ecosystem entails a number of advantages: not only because all stakeholders (even those who would not independently have the operational capacity and resources to build a sufficiently large database on their own) have the possibility to access the data and the benefits that their aggregation can provide, thus constituting an element of democratisation of the system; on the other hand, participation in a data ecosystem promises a number of benefits that extend from brands to individual citizens, who can enjoy commercial offers that are more personalised and adherent to their needs (Olivera et al., 2019).

A further characteristic element of digital marketing is its dynamic nature: this is a feature that depends largely on the extremely pervasive use of technology, which is also in perpetual and continuous evolution. The dynamic nature of digital marketing has at least two implications:

- on the one hand, it requires marketers to be agile and adaptable, confronting technological innovation in a timely manner in order to gain a competitive advantage over competitors; as we have seen, the evolution of digital marketing has been inextricably linked to rapid advances in technology, particularly the rise of mobile devices, social media platforms and cloud computing, which have transformed the way consumers interact with brands and make purchasing decisions. This intricate ecosystem requires a deep understanding of digital channels, platforms and tools, as well as the ability to seamlessly integrate them into a cohesive marketing strategy (Tadimarri et al., 2024)
- on the other, it implies a structure capable of constantly monitoring and modifying campaigns to optimise performance and maximise return on investment; in digital marketing, the interactive nature of digital channels and the exponential increase in digital information consumption facilitates direct engagement with customers,

allowing marketers to build relationships, gather feedback and foster brand loyalty in ways that were previously impossible

This second aspect is particularly important: unlike traditional marketing, which is often based on broad generalisations and assumptions, digital marketing exploits information from large amounts of analytics and performs real-time analysis to understand customer behaviour, preferences and needs, using tools that allow for personalisation of messages and high relevance to consumer interests. These tools, moreover, enable personalised and continuous communication channels with consumers, helping brands to create stable relationships over time and thus secure customer loyalty (Zaman, 2022).

The literature review revealed four main strengths that characterise digital marketing over traditional marketing:

- affordability: the fact that access, analysis and processing of data takes place thanks to the capacity of the network and at a low cost has made it possible for even smaller realities to enjoy benefits that would otherwise only be the prerogative of the big marketing players;
- Ease of monitoring: all responses, reactions and results of interactions between the company and consumers are tracked in real time and form the basis for further evaluation in real time, thus enabling corrections to be made to communication campaigns. This has a twofold advantage: on the one hand, the brand saves time and resources needed to carry out surveys and research on customer behaviour; on the other hand, it enables recovery operations and realignment of the campaign when inefficiencies occur in the effectiveness of the message conveyed;
- allows for a broadening of the reach of campaigns: the number of people connected to the net is growing and constantly expanding. This makes it possible for companies to increase their digital presence and nurture scalable marketing strategies, potentially taking them global with relatively low costs. This makes otherwise inaccessible contexts available, reaching a global audience and opening up new markets and growth opportunities;
- higher conversion rate: studies have shown that digital campaigns have a higher conversion rate: this would be 12.5% (compared to 8% for traditional marketing).

This is largely due to the higher customisation capacity of traditional marketing campaigns compared to digital ones;

- Brand Loyalty: digital marketing has the potential to respond quickly to users, not only helping them in their choice, but also following them up afterwards and resolving critical issues (Erwanda & Doli, 2024).

Digital marketing has fundamentally reshaped the marketing paradigm through the strategic use of the Internet and related digital technologies. Its development testifies, also from an experimental point of view, how the sector is particularly sensitive to incorporating technological innovations into customer communication processes. Companies in the marketing sector are constantly engaged in identifying strategies for communication and conveying promotional messages through the most innovative technologies (IoT, Artificial Intelligence, Blockchain, etc.). The digital transformation in marketing thus appears to be checking every possibility of development, combining the increasing availability of data with advances in AI. These two aspects (information available in relevant quantity, quality and consistency; speed of computation allowing its analysis essentially in real time) has led to the development of a data-driven approach to marketing that represents a true revolution (Pascucci et al., 2023).

Digital marketing has become a critical growth factor not only for brands: the shift to digital channels and platforms has radically changed the behaviour of consumers, who increasingly rely on online resources, social networks and mobile devices to research products, compare prices and make purchasing decisions.

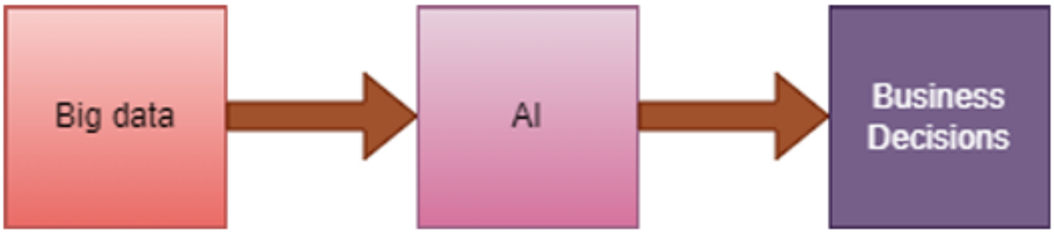
1.2 Evolution of AI in business

The integration of artificial intelligence (AI) into contemporary business contexts represents one of the most significant structural changes of the last decades, taking the form of a true paradigm shift in global business models. From a first phase, in which knowledge was de facto entrusted to rule-based systems and the development of algorithms was guided by human experts, there has been a shift to a phase in which algorithms have become increasingly refined and high-performing and capable of automatic learning. These evolutionary steps are characterised "[...] by an intriguing interplay between human inventiveness and machine capacities" (Bruno, 2024). The development of AI systems has been decisive in contemporary economies: if the availability of data has become a strategic factor, this is also due to the fact that thanks to AI it is now possible to manage, process and make it available very quickly and with a reduced margin of error. In addition, the processing capacity of the machines has also led to greater attention being paid to the quality of the data: in fact, their use depends on their quality and therefore greater care has been devoted to the data production and collection phase.

Today, AI is the protagonist of a process of continuous innovation: we speak of reinforcement learning, which envisages that AI does not learn by trial and error, imitating the learning method of human beings; areas such as unsupervised learning are also being explored, while Generative Adversarial Networks (Goodfellow et al., 2020) and Quantum Computing have left the phase of mere theoretical definition to propose the first (albeit limited, at least as far as Quantum Computing is concerned) concrete applications. These albeit brief considerations, which we will go into in greater detail in the next chapter, help us understand how AI cannot be considered merely as a technological lever, but now takes on the role of a systemic resource, capable of acting as a multiplier of innovation in diversified sectors, from education to industrial automation, from marketing to healthcare.

In today's business landscape, AI emerges as an enabler for improving operational efficiency and optimising decision-making processes through the automatic processing of large volumes of data, predictive analysis and strategic decision support. The fields of use of AI are manifold and are strategic as support for decision-making processes. Organisations are constantly called upon to make decisions, which are often crucial in business life. The problem is how decisions are made. The literature has shown great similarities between humans and machines, but both share the need for information in order to make choices. The problem is that in the age of big data, making rational decisions without incurring cognitive and behavioural biases and at the same time evaluating all possible scenarios is arduous for the human being, who would need extended time to process all available information. AI solves this problem by being able to collect a large amount of data, make predictions and analyse trends much faster and with much lower costs. AI uses data mining to collect and analyse data and thus acts as a medium between data and decision-making (Figure 4).

Figure 4 - AI-Based model for decision making



Source: Prasanth et al., 2023

The ability of AI algorithms to process huge amounts of data also reduces ambiguity and information complexity, enabling companies to react more readily to changes in the competitive environment. As a result, the adaptive capacity of organisations is enhanced, and business agility - understood as the ability to quickly adapt to new market conditions - becomes an essential strategic asset. In fact

> "The advent of Artificial Intelligence (AI) has significantly disrupted traditional decision-making models in business, leading to a paradigm shift in how strategies are formulated and executed. This disruption is not merely a technological advancement but a fundamental change in organisations' cognitive and strategic processes." (Kaggwa et al., 2024, p. 425).

The adoption of AI does not only require the mere implementation of technological tools within companies: it implies a profound organisational transformation of processes (decision-making, management and operational) and even of leadership styles. Companies must reassess their decision-making structures, reorganise internal processes and promote a corporate culture oriented towards continuous learning and innovation. Fundamental are personnel reskilling and upskilling policies, which must accompany AI integration to avoid professional obsolescence phenomena and to ensure effective human-machine synergy.

There are numerous sectoral applications of AI, which by its very nature lends itself to be particularly 'plastic' with respect to different needs in multiple fields of human resources management, where AI has taken a central role in personnel selection, evaluation and training processes, thanks to AI-based *Applicant Tracking* Systems can automatically analyse resumes and cover letters to identify the most qualified candidates based on specific criteria (Jatobá et al., 2019). These intelligent systems are used to analyse candidate profiles, conduct automatic screenings and assess the compatibility between candidates and required roles, based on behavioural parameters, cognitive tests and soft skills. HR analytics also makes it possible to predict phenomena such as employee turnover or dissatisfaction, facilitating more effective retention strategies. Furthermore, AI can be used to conduct automated *video interviews*, evaluating candidates' responses and body language to predict their ability to fit into the new professional environment. It is precisely predictive analysis that appears to be one of the most promising fields in the application of the use of AI in human resources management: it allows for future scenarios to be drawn up with a high margin of accuracy, for instance by predicting the real hourly cost of human resources (Abbracciavento et al., 2020) or by analysing the risk of employees leaving through demographic data, performance, and feedback, allowing for proactive intervention to improve staff *retention*. The advantage that AI offers is related not only to the possibility of taking into account a greater number of parameters than a human being can, but also to evaluate them objectively, reducing the weight of cognitive and relational biases in all phases of human resources management, starting from recruitment, using a series of appropriate tools, as shown in Table 1 (Madanchian, 2024):

Tab. 1 - Impact of AI on HR, with mention of mail AI Applications

Key Aspect	Description	Impact on Recruitment	AI Application
Efficiency	The time spent on administrative duties, such as resume screening and initial assessments, is significantly reduced by AI. This allows recruiters to focus more on engaging with candidates and making well-informed decisions.	More time allocated for candidate engagement and strategic decision-making.	Resume screening tools, chatbots
Quality of Hire	AI can identify candidates who are the best fit for specific roles based on their abilities and experiences by utilizing data analytics, leading to higher-quality hires.	Improved matching between candidate skills and job requirements, reducing turnover rates.	Predictive analytics, skill matching
Improved Candidate Experience	AI applications improve communication with candidates by providing timely updates and feedback throughout the recruitment process, enhancing the overall experience.	Enhances transparency, increases candidate satisfaction, and strengthens employer brand.	Automated communication platforms

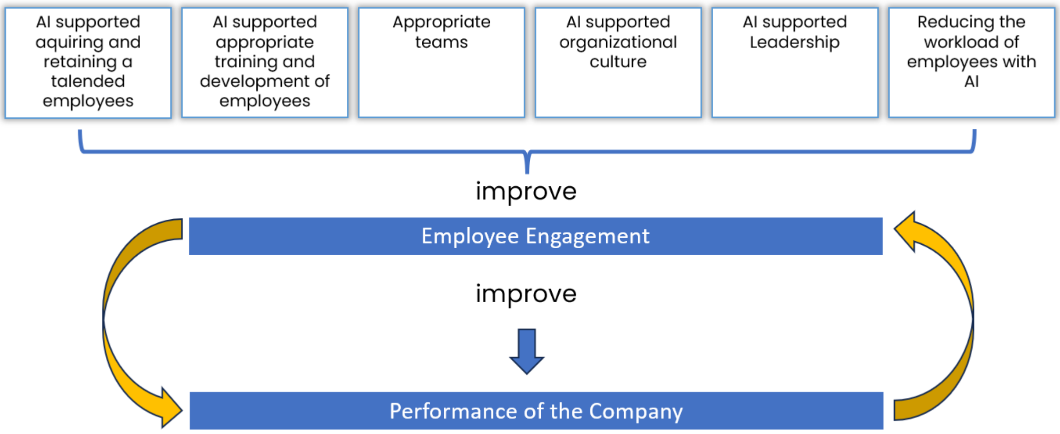
Source: Madanchian, 2024

AI can improve the allocation of personnel to the most suitable roles, taking into account employees' competencies, preferences and soft skills. This not only increases employee satisfaction, but also overall organisational efficiency by intervening in what the literature calls 'talent management': this is a particular predisposition of companies that tends to promote the development of human resources within the company and enhance their talents by developing a sense of belonging and participation in decision-making processes. This is an element that is considered particularly important for companies today: to survive in increasingly complex and competitive markets, in fact, it is necessary to preserve the most important strategic resource. Personnel, especially when they are trained and competent, are difficult to find on the market and keeping them in the company has always been a primary objective of every organisation. AI can intervene in this area with various solutions, ranging from the creation of teams that balance the relational and operational potential and limitations of each member to the creation of an innovation-oriented corporate culture (Rožman et al., 2022).

Incorporating AI into business processes means that talent in the organisation can be freed up by reducing administrative and repetitive tasks in HR: these tasks can be automated through AI and allow top management to use professional resources on strategic and high value-added activities.

Artificial intelligence is also being used to customise training paths, through adaptive learning platforms that modulate content according to prior skills and achievements, facilitating more targeted and efficient learning. These tools transform the HR function from operational to strategic, enabling smarter human capital management. This is related to employee engagement and company performance: issues in which AI plays a key role (Rožman et al., 2022). Figure 5 shows how AI can simultaneously contribute in many fields to improving corporate performance on the one hand, and the emotional, cognitive and behavioural engagement of employees on the other:

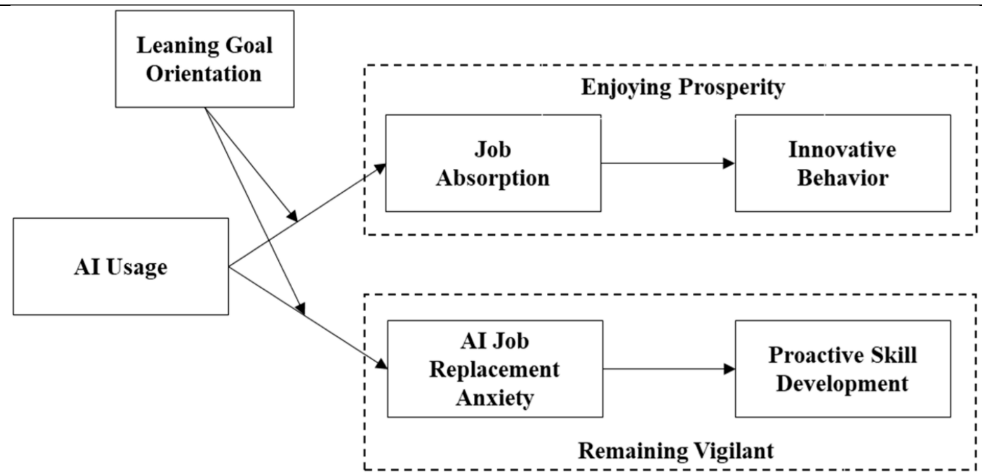
Figure 5 - Embedded aspects of AI in HR



Source: personal draft based on Rožman 2022, p. 7

From an operational point of view, the integration of AI in businesses enables a more efficient management of resources, reducing execution times and cutting costs related to human error, information redundancy or failure to optimise flows, and above all improving interactions between personnel, technology, organisational structure and environment. These are crucial aspects for the competitiveness of companies: the benefits that AI can bring are therefore penetrating organisational structures more and more deeply, to the point that the global adoption rate has more than doubled between 2017 and 2023, from 20% to 55%. The introduction of AI into the operational processes of companies entails significant consequences that the literature has highlighted. In addition to the advantages and positive aspects, however, there are also problems that it is important not to underestimate. While the introduction of AI has positive effects in the performance of highly repetitive and routine tasks because it frees up employees' mental resources and redirects their cognitive abilities and creativity towards different and more challenging goals, it has also been shown that the pervasiveness of AI limits the involvement of employees in decision-making processes and has repercussions on their social self-perception and sense of belonging to the organisation (Qian et al., 2025). This results in two distinct visions: one that emphasises the positivity of the use of AI (termed "Enjoying Prosperity"), the other that views it with apprehension, leading to the phenomenon termed "replacement anxiety" and which is all the higher in contexts where objectives are more learning-oriented ("Remaining Vigilant") (Figure 6)

Figure 6 - Different approaches to AI in HR management

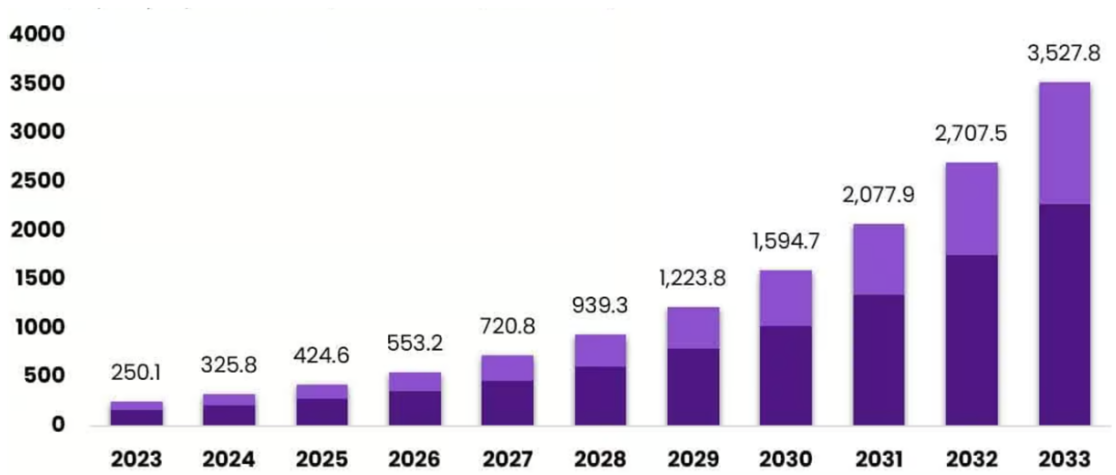


Source: Qian et al, 2025

The positive effects of this approach have been measured in the literature, which has shown that the use of AI has positive effects on employees' innovative behaviour and their ability to develop proactive skills.

The growing confidence in the transformative potential of artificial intelligence is not only about efficiency, but is also reflected in macroeconomic dynamics: according to some estimates, the global AI market (which stood at around 250 billion dollars in 2022), could grow to more than 3,500 billion dollars in 2033, with an annual growth rate of more than 30 per cent, thus contributing up to 15 per cent of the global Gross Domestic Product (GDP) by 2030 (Figure 7). This figure underlines the enormous potential of AI as a driver of economic growth and systemic innovation (Market Us, 2024).

Figure 7 - Global Artificial Intelligence Market



Source: Market Us, 2024 https://market.us/report/artificial-intelligence-market/

The concrete applications of AI in business are many and increasingly sophisticated. Among the most widespread are:

- **Intelligent Process Automation** (IPA), which combines robotics and machine learning to perform repetitive and complex tasks, particularly with regard to efficient supply chain management;

- **predictive analytics**, which makes it possible to forecast market trends or consumer behaviour through the processing of historical and real-time data;
- **dynamic personalisation of the customer experience**, which dynamically adapts products, services and communication to individual user preferences (Dash et al., 2019).

These technologies not only increase productivity, but also foster strategic differentiation. Pioneering companies are investing in AI to consolidate their competitive advantage, strengthen customer loyalty and generate new sources of value. In the financial sector, for example, AI is being used for fraud detection, proactive customer relationship management and credit risk optimisation: in this sector in particular, the ability of AI applications to manage data from different sources and aggregate them according to innovative methodologies allows (also thanks to predictive algorithms) to operate transparently and to balance the interests of credit institutions and customers, contributing to a more open and proactive relationship (Locatelli et al., 2018).

AI is also being used profitably in the health sector: the literature points to approaches that claim that here too AI can have a revolutionary development due to its ability to increase the accuracy of diagnoses, improve the personalisation of treatments, and predictive medicine. Considering that the healthcare field is one of those in which scientific and technological innovation is most involved (think of the development of robotic surgery or telemedicine), it is conceivable that AI could play an even more significant role in the near future, improving patient outcomes and supporting scientific research (Bhagat et al., 2024).

The adoption of AI is also transforming areas such as agriculture, where it is used to monitor environmental conditions and optimise resource use, and cybersecurity, thanks to threat detection systems based on machine learning and adaptive mechanisms (Meshram et al., 2025). These applications do not merely replace human capabilities, but enhance them, freeing up time for more strategic and creative tasks, with clear benefits on the quality of management and the innovative potential of businesses.

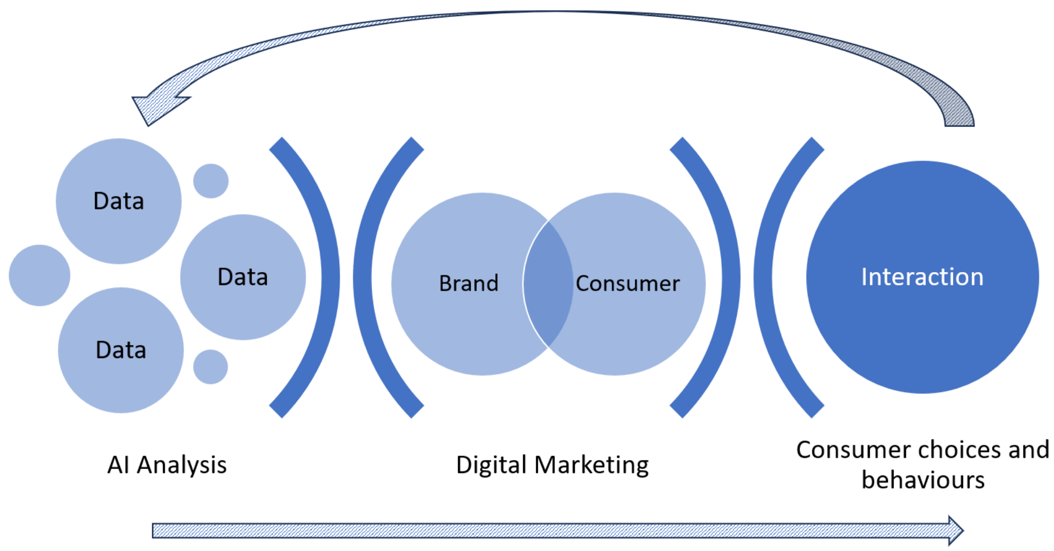
Such integration leads organisations to develop new forms of strategic intelligence, in which automated data analysis fuels more informed and timely decisions. The deployment of AI technologies in business planning, market intelligence and business consulting functions - including strategy definition, the study of markets and process re-engineering - is an example of how the added value of AI now cuts across all managerial activities.

The evolution of AI in the business context should not be read as a mere technological upgrade, but as a process of systemic transformation involving the organisational, cultural and economic dimensions of enterprises. It opens up unprecedented scenarios for competitiveness and efficiency, but requires responsible governance and a continuous capacity to adapt and learn.

1.3 The impact of AI on marketing processes: from personalisation to automation

The spread of the intranet and the increasing use of mobile devices has produced a revolution in marketing. Access to marketing resources through digital has also become possible for small and medium-sized enterprises, which have been able to compete with large companies on a more equal footing. Moreover, thanks to continuous innovations in data management and digital marketing applications, the ability to collect, analyse and target data has been amplified, leading consumers to use the information for their purchasing choices. In fact, a circular process has been triggered: (i) the analyses of the available data enable companies to use marketing tools to improve their business propositions and after-sales service, reaching more potential customers and building consumer loyalty; (ii) consumers increasingly use digital tools for their purchasing decisions, to stay in touch with brands and to express ratings on the quality of the product, service or support; iii) their online interactions and purchase choices further feed the amount of data that, thanks to the potential of AI, return to the availability of companies, which can refine, correct and improve their product and service offerings (Fig. 8).

Fig. 8 - Digital Marketing process



Source: personal draft

In this process, AI is decisive and performant, to the extent that the last decade has seen a gradual but inexorable process of replacing traditional marketing strategies with digital marketing, which is now prevalent. This has led to the need for companies to develop increasingly in-depth skills in digital marketing (Liu, 2022).

The integration of artificial intelligence into marketing strategies has marked an epoch-making turning point, radically transforming the way companies interact with consumers to promote brand growth and build customer loyalty by improving their engagement with the brand (Tadimarri et al., 2024). Artificial intelligence has emerged as a strategic tool, fundamental to enabling brands to connect, interact and engage with their target audience. This is an inescapable need in the contemporary marketing world, in which the consumer is no longer just the object of promotional campaigns, but chooses in an increasingly conscious and informed manner. This capacity for interaction is facilitated by AI: which, by being able to analyse large datasets in real time, enables brands to obtain information on consumer behaviour, preferences and trends and activate two-way communication, which is much more effective than traditional campaigns, based on predefined targeting from which it was difficult, if not impossible, to obtain feedback capable of supporting the company's communication effort. Instead, the information returned through customer-brand interactions conveyed by the network and analysed thanks to AI is valuable because it can be used to fine-tune personalised marketing campaigns. Personalisation, a key component of modern marketing, is greatly enhanced by artificial intelligence algorithms that propose content and recommendations to potential customers that depend on their own information and individual user data, appropriately aggregated. This results in a better consumer experience, leading to greater customer satisfaction and engagement. Brand leaders show the transformative potential of artificial intelligence integrations in marketing, which is not just a technological tool, but a real catalyst for change in the marketing sector, capable of opening new frontiers and redefining the rules of the game. The literature has identified six key themes, which are in turn divided into three distinct disciplinary areas (Fig. 8):

- **Analitycal Marketing capabilitis** area: Develops the ability to deepen analysis on the basis of data provided by users, with the aim of understanding their preferences and elaborating a coherent and customised offer. It is developed through two main themes:
  - *AI-driven customer insights*: data is used to predict future user behaviour, enabling marketers to anticipate consumer needs. This is an innovative approach because it is not reactive to personal needs, but based on data from previous experiences. AI, in this respect, is decisive in the development of CRMs, which in turn are instrumental in the predictive analysis of consumer behaviour;
  - *measuring marketing performance*: AI offers companies the ability to accurately measure the performance of marketing campaigns. This is made possible by the ability to analyse the data that is generated from consumer interactions. AI-powered analytics tools are used to perform this analysis, using algorithms capable of interpreting significant amounts of data and understanding performance results. AI also allows performance to be tracked in real time, increasing the possibility of understanding the effectiveness of marketing campaigns;
- **Technological marketing capabilities** area: to this area belong the insights related to the interaction of different technologies in digital marketing processes. The main technologies used (programmed advertising, chatbots, virtual assistants, etc.) are particularly effective tools that show a high capacity to integrate with marketing strategies. However, they also entail a number of ethical issues,

particularly related to data security and the integrity and confidentiality of the information that consumers 'give' to brands. From a technological point of view, it is interesting to point out that algorithms are not exempt from cognitive biases either: research in this field has indeed highlighted how they can reinforce prejudices, leading to potentially harmful actions towards certain social groups (Ferrer et al., 2021).

- *Automated marking strategies*. Automation makes it possible to optimise the execution of particularly demanding analysis tasks, ensuring the personalisation of communication and identifying a precise target group of potential customers. In the field of automated marketing strategies, two techniques have emerged: programmatic advertising and chatbots. Programmatic advertising is a tool that is growing steadily (+13% between 2022 and 2023) and PA budgets are estimated to grow by 300 billions of dollars in the coming years. This type of campaign uses AI algorithms to reach specific targets through different digital channels (video ads, TV advertising, digital out-of-home etc.), combined to ensure not only that the message is perceived, but also that it produces a high conversion rate from potential consumers. Chatbots and virtual assistants represent a profound change in customer interactions. Chatbots are artificial intelligence engines that execute sequential dialogues to achieve an interaction with humans. They have developed since the mid-2010s, offering companies a new possibility to interact with customers. Over time, they have evolved into intelligent assistants (think of Amazon's Alexa and Apple's Siri). They are powerful tools, capable of communicating with humans using natural language and for this reason particularly friendly in interactions, also being able to recognise, simulate and react to emotions, a capability that is rapidly improving;
- *Ethical implications*. AI offers performance based on the analysis of large datasets from a variety of ecosystems (from healthcare to finance; from education to justice). Algorithms offer a high degree of automation capacity, which however also entails ethical challenges related to privacy and data security, with implications that are not only technological, but

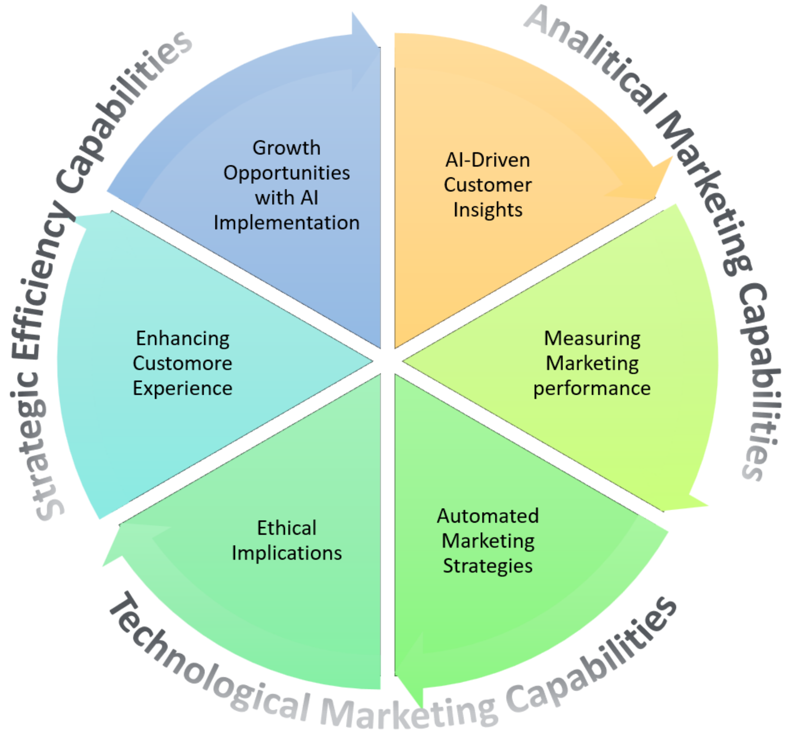
also extend to the social, cultural and psychological level. Privacy (i.e. ensuring that the data provided by users are treated with due respect for the confidentiality due to 'sensitive' information, relating to sexual, political, religious orientations and people's fragile conditions) and data security are the key elements in this area;

- **Strategic efficiency capabilities** area. AI improves the customer experience, but it also implies the need to strike a balance between the enormous innovative potential of AI and its implementation in practice. This involves balancing the possibilities of offering quality services to customers and promoting business growth through the potential of AI.
    - *Enhancing customer experience*. AI makes it possible to provide immediate assistance to customers, ensuring that requests can be handled effectively and quickly; it selects and offers information relevant to users' needs and propensities; and it ensures interactions that simplify purchase choices and quality services. For the most part, AI interacts with physical space (through geo-localisation), fostering interactions between the virtual and real dimensions: think, for instance, of the strategies adopted by many brands, which involve choosing goods online and picking them up in a physical point of sale. In the area of improving the user experience, however, there is still only partially explored ground. This is the relationship between users and virtual environments (extended or augmented reality, spatial computing, metaverse). Many leading companies in their respective fields are developing Extended Reality (XR) to offer their customers new stimulating experiences. This is a rising market based on generative AI, which creates virtual environments and language models where customers and brands can interact. The integration of AI and XR has applications in several sectors: the most immediate (and in part already the subject of interesting practical developments) is gaming, where the technology can not only create realistic scenarios, which are also rendered as such by immersive visors; AI can in fact adapt virtual worlds to human interaction, giving even more of a feeling of being immersed in a real world. In the field of commerce, the combination of

Extended Reality and AI is used to create virtual shop windows. The combined application of these two technologies presents a number of obstacles: there are technical and accessibility issues, legal and health concerns (prolonged exposure to Extended Reality can lead to reality perception problems and physical problems such as eye fatigue and motion sickness). Not least, the development of applications driven by these technologies is currently extremely expensive. More current, however, is the state of the art concerning real-time assistance: improving the customer experience also depends on the speed of response to requests for assistance, and AI is already a valuable ally for companies in this area;

- *Growth opportunities with AI implementation*. In an economy increasingly dominated by digital and data, AI can effectively address the growth and competitiveness needs of companies. However, it is essential to assess the application of AI in marketing through the ratio of costs to benefits. Besides the undoubted benefits, there are also limitations and risks that cannot be underestimated. In fact, it must be considered that the benefits are not acquired once and for all: the speed with which the technology advances means that the advantages for early adopters are not long-lasting, both because competitors will sooner or later adopt similar strategies, closing the initial gap; and because the introduction of new tools and new strategies risks quickly making those adopted obsolete. More significant seems instead the adoption of a long-term strategy: it requires, however, to incorporate the interaction between man and machine as a fundamental element in business and marketing processes, avoiding the risk of relying exclusively on technology, assigning it a saving role; furthermore, a strategy that differentiates interaction with customers on the basis of profitability evaluations seems more effective (Kumar et al., 2024).

Fig. 9 - Marketing areas and themes



Source: personal Draft based on Kumar, 2024

Analyses based on artificial intelligence enable brands to anticipate market changes and optimise strategies precisely because of their ability to process increasingly important data sets. The wealth of information that brands have at their disposal enables them to proactively adapt strategies, anticipate emerging trends and stay one step ahead of the competition, increasing their competitiveness. Within the artificial intelligence landscape, predictive analytics emerges as a powerful tool for extracting value from data and anticipating future trends, such as those related to Extended Reality we mentioned (Dobosevych, 2023). AI leverages advanced algorithms, such as machine learning and neural networks, to analyse historical data and identify meaningful patterns, enabling the prediction of future outcomes and behaviour, supporting companies in making more informed and proactive decisions, optimising marketing strategies, improving operational efficiency and mitigating risks. It enables real-time adjustments, ensuring that marketing strategies are agile and aligned with changing market dynamics. AI-based analytics provides brands with the foresight to make data-driven decisions, improving relevance and effectiveness in a dynamic business landscape. While artificial intelligence has immense potential, its responsible and ethical use is essential to maintain consumer trust and ensure transparency in data usage and decision-making processes.

With the ability to identify patterns and correlations that would escape human analysis, AI enables marketers to segment audiences more precisely, personalise marketing campaigns and optimise the allocation of advertising budgets). The implementation of AI-based automation systems allows for the optimisation of repetitive and standardisable tasks, such as the sending of personalised emails, content management on social media and basic customer support. This frees up valuable time and resources for marketing teams to focus on more strategic and creative activities, such as developing new campaigns, analysing results and product innovation. Artificial intelligence also makes it possible to improve the customer experience through intelligent chatbots, virtual assistants and customised recommendation systems. These tools make it possible to provide fast, efficient and personalised assistance to customers, improving customer satisfaction and loyalty.

The evolution of predictive marketing has accentuated the understanding of consumer decision-making, while artificial intelligence enables many companies to predict big consumer data to meet customer expectations and provide personalised products and services (Zaman, 2022). The implementation of artificial intelligence in marketing is on the rise as it offers opportunities such as personalisation and an enhanced customer experience, ensuring the company increases its performance and return on investment.

2. FROM TRADITIONAL AI TO GENERATIVE AI AND AI AGENTS

2.1 Artificial intelligence: from machine learning to deep learning

Technological evolution is thus a key driver of AI applied to business. It began its journey from the development of Machine Learning (ML), which has its basis in computational statistics: with this discipline it shares the goal of processing, i.e., making predictions using the large (and growing) computational power that modern computers can provide. At the core of Machine Learning is the ability of systems to autonomously identify patterns and make decisions, learning from data and minimizing human intervention.

The origins of ML date back to the late 1950s. Arthur Samuel, who is credited with coining the term "machine-learning," was the first to glimpse the possibility that

> "A computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program" (Samuel, 1959).

This insight of the American scientist highlights some elements that seem particularly important and that ML applications seem to confirm both on the level of theoretical development and in the area of its operational applications. In particular:

- the ability of computers to perform specific tasks without being programmed to do so: basically, computers learn from previous processing and patterns to deliver the expected results, making decisions that are reproducible and the more reliable the more complex the machine's ability to apply the algorithms defined by the developer;
- the possibility that a computer capable of learning from experience can avoid the complex programming phase, thus minimizing human intervention. The developer operating in a Machine Learning context has a very different task to the writing of a traditional program: instead of writing the code through which the machine is told what sequences of operations are to be performed, he defines a more or less complex set of data (*dataset*) on which the computer, through

algorithms, is called upon to develop its own logic to perform the required function, activity or tasks.

For Machine Learning, data, their consistency and integrity are crucial: a number of operational choices depend on their availability. The great emphasis placed on data brings this approach closer to Data Mining: the latter, however, focuses mainly on exploratory analysis (a function that can, after all, also be performed through Machine Learning techniques); with Machine Learning, on the other hand, classification, regression and data interpretation models can be implemented based on characteristics already known and related to a more or less extensive sample of the data itself, on which system training is exercised. The machine "learns" to recognize features and relationships and to build complex models, which it adapts over time because it can improve its performance improve after a task is performed, even when this task has produced an error (Buczak & Guven, 2016).

ML has found further development in Deep Learning (DL) and Neural Networks (NN). Deep learning has developed over the past decades and has characterized the interest of researchers and companies, as evidenced by increasing investments. Deep Learning is one of the subsets that make up the Machine Learning galaxy. Although-even on a terminological level-this new direction in Artificial Intelligence is rather recent, again its foundations have roots that go back as far as the 1940s and the first studies in cybernetics, when people had begun to think, at least on a theoretical level, about the possibility of making algorithms capable of replicating the functioning of biological neural networks.

In the late 1950s, psychologist Rosemblatt had outlined an absolutely fascinating and innovative research framework:

> "*Since the advent of electronic computers and odern servo systems, an increasing amoung of attention has been focused on the feasibility of contruscting a device possessing such human-like functions as perception, recognition, concept formation, and the ability to generalize from experience*" (Rosemblatt, 1957).

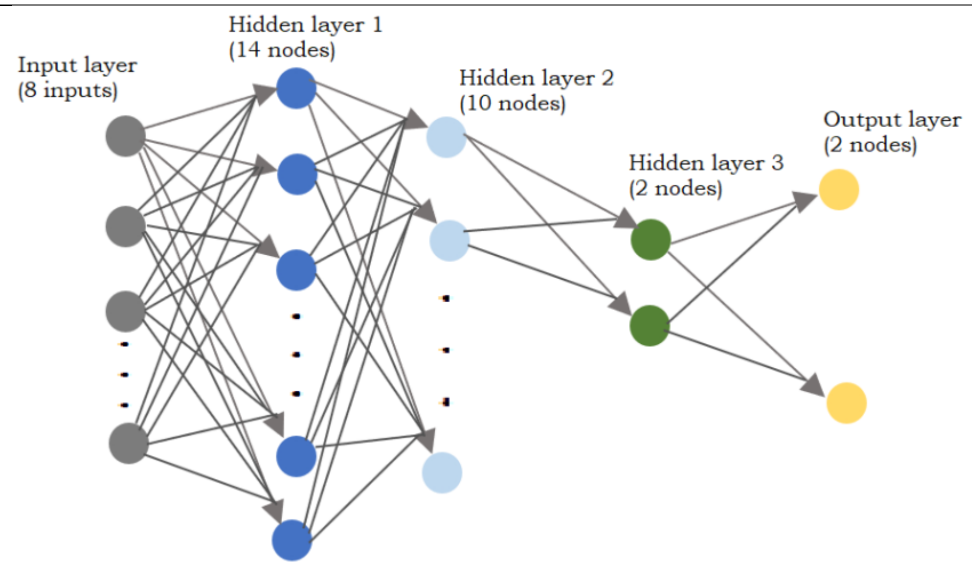
Rosemblatt's project had actually seen the construction of Perceptron, an artificial neural network that was powered by a huge computer occupying the size of a room. Perceptron's functions were limited and were entirely devoted to dividing a set of tiles into two distinct groups, discriminating those on the right from those on the left. Nevertheless, even in the simplicity of the algorithm, this first step fueled a certain enthusiasm, giving a decisive boost to cybernetics studies. The scientific community remained substantially divided: while on the one hand from the columns of the New York Times some scholars went so far as to "predict" the start of robotics, many initially remained skeptical about the real capabilities of a computer to perform tasks autonomously and such as to replicate the behavior of the human brain (Olazarn, 1993).

As much as the algorithms that underpinned Perceptron were mainly used to understand and study the structure of human neuronal connessionins rather than to carry out predictions, Rosemblatt's pioneering work had the merit of opening up an important theoretical elaboration. This also struggled for a long time to find feedback in operational practice because of technological limitations that prevented the development of systems capable of applying and testing more advanced systems than Perceptron. This was a limitation that would last for a long time, although already during the 1980s LeCun and Hinton had added a fundamental step to the theory, namely, the possibility of developing programs capable of enabling artificial neural networks to correct errors autonomously, also incorporating into the field of cybernetics and artificial neural networks some concepts related to *backpropagation* (*Backpropagation*) of error that characterizes the human brain (LeCun, 1988). But still technology was not able to properly support the operational development of such hypotheses.

In 2012, finally, George E. Hinton was able to present an experiment (the *ImageNet contest*) in which a computer proved capable of recognizing humans and animals by comparison with millions of images stored in it and without the need for coded human intervention (Lachure et al., 2023). The success of the trial was replicated only three years later and with even more surprising results: in fact, in 2015 the result of a trial was presented that had succeeded in reaching 152 levels of abstraction: a result more than five times higher than the previous standard. It means that the machine was able to use on the same image and simultaneously 152 algorithms, corresponding to as many simultaneous operations from which the machine drew information to "learn" without the need for human intervention to preset information, tags and other meta-data.

This track outlines, albeit briefly, the introduction of Deep Learning as a promising area of research. The reason why Deep Learning differs from the more general domain of Machine Learning is related to a number of factors, the most significant of which are the possibility of operating in a nonlinear manner, replicating through algorithms biological neural networks. The realization of artificial neural networks is possible because their mathematical models constitute processing nodes that, in analogy with biological networks, are identified as "artificial neurons": these exploit, just like the neural networks of the human or animal brain, the ability to interconnect information. During the 1980s, the stream of cognitive science studies called connectionism had proposed a new approach for software programming: in a neural network, information took place according to the principle of Parallel Distributed Process (PDP): the human brain operation processes information from the various senses in a parallel manner: acquiring the input and distributing it simultaneously to all nodes in the network. Thus, there is no central "memory" in which information is stored and then passed on to the various points: rather, it is immediately distributed across multiple levels and simultaneously; each level reached by the input, in turn, presides over autonomous processing that helps define the output result of the process. The next figure illustrates the classic neural network model:

Fig. 10 - Example of a neural network



Source: Halbouni et al., 2021

This brief description of the process that takes place within a neural network immediately gives an idea of how different their operation is from the processes of classical computer science: these in fact involve the existence of a memory in which data are stored and subsequently processed through a series of sequential calculations. Their transmission to the various points in the network is also done sequentially, with a burden on resources, accuracy and time to get the output processed (Halbouni et al., 2021).

Artificial neural networks consist of artificial neurons, also called ‘processing units,’ connected to each other in a manner similar to the connections in biological neural networks. As we have seen, the analogy between artificial neural networks and biological ones does not stop at morphological aspects alone, but also at the ways in which they perform computations aimed at the task that is required of them. These not only take place through a series of parallel elaborations, but also make it possible for the machine to autonomously and constantly feed its own computational capabilities and accuracy from the elaborations that are performed from time to time. It is this second characteristic that justifies the name of this approach: Deep Learning (translatable as "deep learning") indicates precisely the ability of artificial neural networks to learn without human intervention, as happens with the Machine Learning processes we saw in the previous chapter.

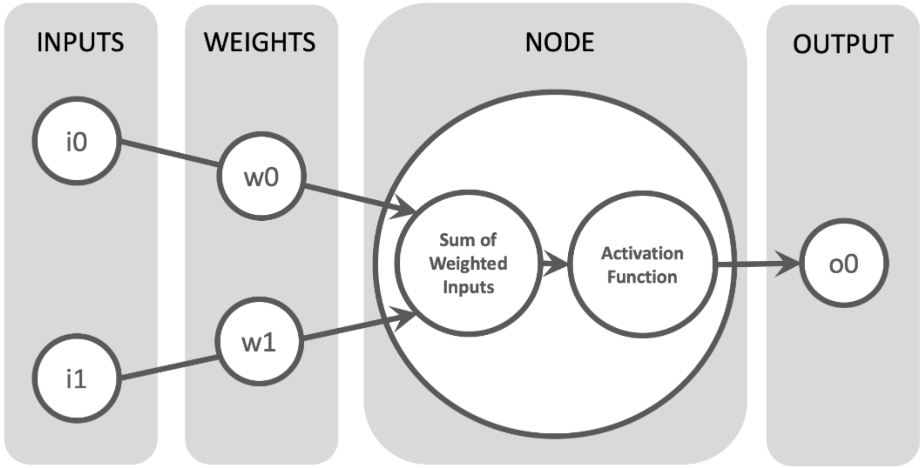
Between the input and output layers there may be one or more intermediate layers (Hidden layers), composed of different processing units (or, as we have seen, artificial neurons). Neural networks are called "feedforward" when they are formed by an architecture structured in several layers of processing units; furthermore, a feedforward network is called "fully-connected" when each processing unit is connected with all the processing units in the next layer.

Compared with ordinary computer networks, deep neural networks use more intermediate layers (called Hidden Layers) to build more levels of abstraction. Each intermediate layer, in fact, has the ability to recognize specific patterns of information and it is evident that the more layers, the more structural details that can be related to each other from the input data in order to obtain a solid output data. The result is that the greater the number of intermediate layers, the greater will be the abstraction capacity of the deep neural network: this will lead to an advantage because complex problems in task recognition can be solved, since at each intermediate layer additional, in-depth information is identified and added to complete the analysis (LeCun et al., 2015).

Among the key features of Deep Learning is the system's ability to learn automatically from data rather than involving a programming step: this feature (which is shared with many applications of Artificial Intelligence, including Machine Learning) is amplified in the case of Deep Learning.

As is the case with Machine Learning, Deep Learning also requires a training process, which is essential for the model to learn to define internal representations of the data that are based on relevant information related to the assigned task: in an image recognition model, for example, the learning phase will aim to facilitate the identification of the fundamental characteristics of the images that are submitted to it (edges, shapes, patterns etc.). For the effectiveness of the model's interpretation, it will need to have – in the training phase – a significant amount of data, so as to extract and acquire as many features as possible, which will then be compared with the actual data during the operational phase. The automaticity of learning comes from the model's ability to learn through optimizing the weights of neural connections. In fact, each piece of information learned in a Deep Learning model traverses the artificial neural network by "touching" one or more nodes: the neural connection weight expresses the relative importance of each connection in the network transmission. When a signal is propagated through a connection from one neuron to another, the connection weight will thus be decisive in defining the importance of that signal.

Figure 11 - Diagram of a neural network with evidence of the weight of connections



Source: West M. (s.d.)

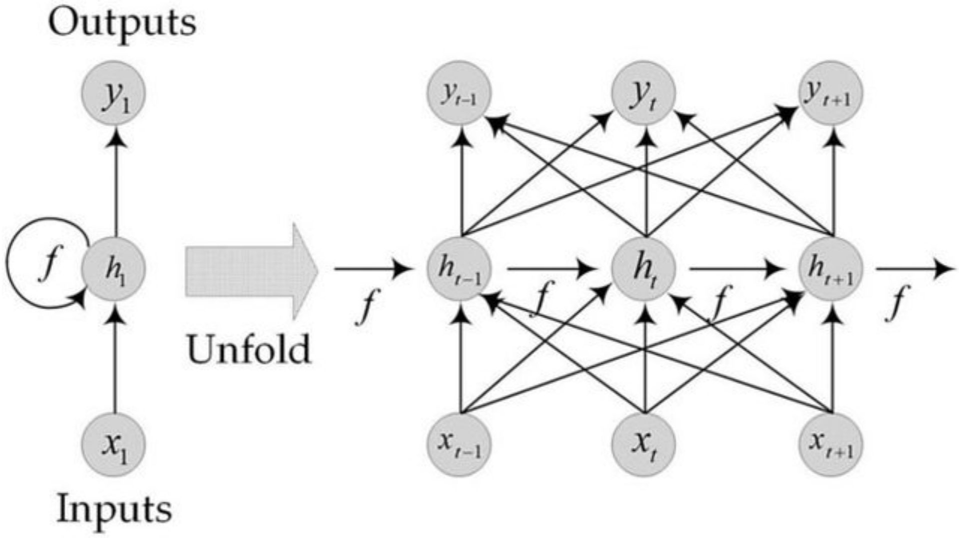
During the training phase, connection weights are adjusted to achieve optimal model performance, finalizing the computational capability to provide accurate predictions and reduce the error between model predictions and training data. This is done through the model's ability to detect and capture relevant features contained in the data (edges of an image; key words in a text etc.). In order to optimize and maximize the benefits of the model, the algorithm that calculates the update of the weights of the neural connections is particularly important so as to avoid the error between the model prediction and the training data. In this way, it is possible to avoid backpropagation of the error (backpropagation), an aspect that we had already seen to be crucial in the early theoretical studies of artificial neural networks in the last century: technically this is achieved by making sure to reduce the error by sending the information to the input layer and identifying it as an error, automatically updating the weights of the neural connections (Mishra et al., 2021).

Rosemblatt already had mentioned how one of the fundamental goals of artificial neural networks was "[...] *the ability to generalize from experience*." This characteristic is also one of the major features of Deep Learning: in essence, the ability to generalize consists of the model's ability to apply what it has learned from the training data to new data that will be submitted to it that were not previously known. This is a feature and capability that is critical for the model to be effective not only on training data, but also on real-world data. For the model to work well on real-world data as well, several steps are required during the training process and-as we have already seen in Machine Learning processes-it also involves avoiding the risk of overfitting, that is, overfitting the model to the training data.

In addition to the Feedforward neural network, which is the classical representation of a neural network, there are several types of neural networks that are used in Deep Learning processes. Among them, the most significant are:

- **Recurrent Neural Networks (RNNs)**. The recurrent neural network architecture is particularly suitable for managing and evaluating sequential data or data that have a temporal structure: this is the case with texts (for example, for their translation or cataloging), audio files or for managing data organized in more or less complex time series. This type of feature makes RNNs particularly effective in the financial field-where many time series of macroeconomic data must be managed-or for making weather forecasts. RNNs differ from other neural networks (and in particular from what we have called feedforward neural networks) because information is not unidirectional, but flows through a series of recurring cycles or connections. Each cycle stores in memory the information acquired during the previous connection. The basic unit is the so-called "recurrent neuron": it maintains an internal state that is updated during the various cycles. It is widely used the LSTM (Long Short-Term Memory) unit, which has the characteristic of capturing long-term dependencies and limiting the main impact of RNNs, i.e., the disappearance of the gradient: this is a phenomenon in neural networks that records that in the process of back-propagation of the error, the signal decreases exponentially as a function of distance from the final layer (Tadjer et al., 2021). This implies that when very long sequences are processed, the gradient that is used to update the weights becomes very small and thus becomes inefficient. The next figure shows the structure of a RNN:

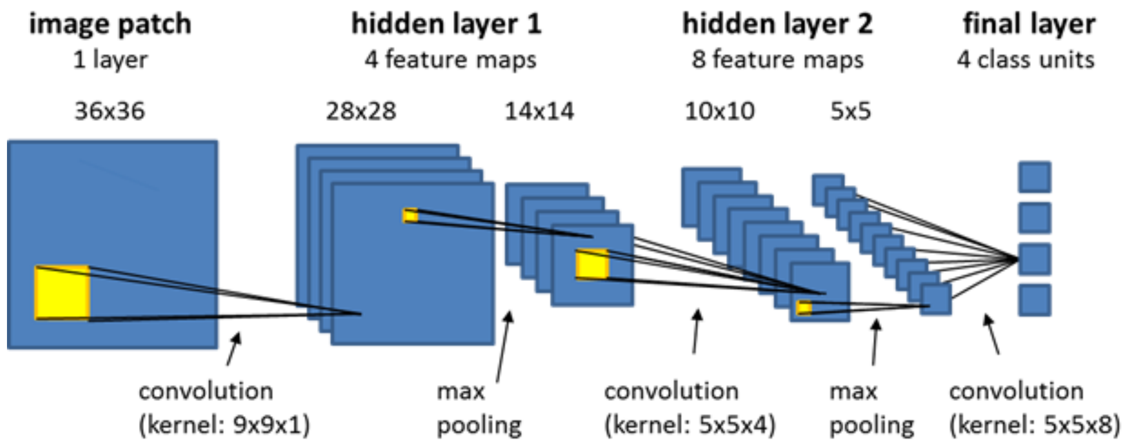
Figure 12 - Structure of a Recurrent Neural Network (RNN).



Source: Tadjer et al., 2021

- **Convolutional Neural Networks (CNNs)**. These are networks that are specifically designed for efficient image recognition and processing, a process in which they are particularly efficient. A CNN operates much like biological processes in the brain. The input layer is the one assigned to accept the source image or data: being particularly suited for use on images, CNNs typically operate on a grid of pixels. The first intermediate layers (Hidden layers) are called convolutional and have the characteristic of using filters (kernels) capable of capturing and defining the input image to derive the most relevant features and the main patterns that compose it. Once this is done on the convolutional layers, the information is transferred to the pooling layers, which reduce the spatial dimension of the representation and extract from it a further selection of features deemed relevant for image recognition. Subsequent intermediate layers are *fully-connected*, as is the case in traditional neural networks: it is in the processes that occur in these layers that the features extracted in the first layers are analyzed and structured to perform recognition, classification or other functions. In a CNN, each artificial neuron is associated with an activation function (ReLU, Rectified Linear Unit): it is in relation to this particular type of operation that nonlinearity of operations is guaranteed, an aspect that amplifies the ability of the neural network to learn from the complex data it has just processed. As in any neural network, each connection between neurons has an associated weight whose parameter is defined in the training phase. However, in CNNs the weights are shared between different regions of the image, allowing the model to recognize similar patterns occurring in different regions of the image-this process underlies the ability of CNNs to more effectively recognize the variation in position of the processed subjects. The next figure summarizes how a CNN works:

Figure 13 - Convolutional neural network (CNN) model.



Source: Trimble eCognition Suite (www.docs.ecognition.com)

Over time, this approach has had further developments due to the growing importance of neural networks and deep learning, enabling AI to analyze huge amounts of data quickly, identify complex patterns, and perform image recognition and natural language processing tasks.

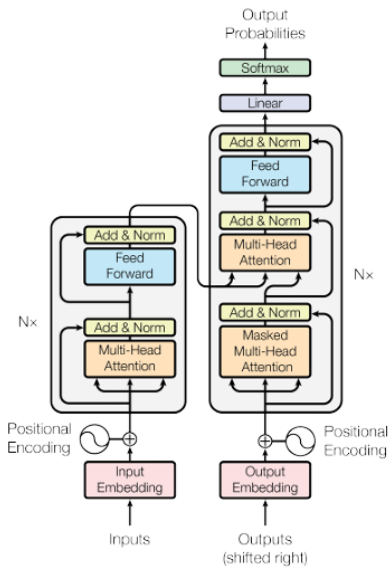
2.2 The Generative AI revolution (linguistic and multimodal models)

The progressive and extraordinary acceleration of AI development has culminated in the advent of generative AI. This is a branch of AI capable of producing original content in the form of text, images, sound, and video and represents a momentous turning point in the relationship between humans and machines. In fact, the development of generative AI implies not only technological progress, but also a cultural, social and economic transformation that could potentially redefine the boundaries of creativity, knowledge and automation.

According to some of the literature, this is a real revolution, underlying which is the development of so-called Large Language Models (LLMs), such as GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers). More recently, these models have been integrated into multimodal architectures capable of simultaneously processing textual, visual, auditory, and, in some cases, sensory inputs. The result is an artificial intelligence that is not only more powerful, but closer to human perceptual and cognitive complexity.

The evolution of language models can be read as a gradual transition from simple statistical prediction of words to simulation of cognitive processes. Early models were based on simple n-grams and recurrent neural networks, which processed words sequentially. The introduction of the Transformer architecture (Vaswani et al., 2017) ushered in a new era in the sequential handling of data (particularly sentences and text documents): it overcomes the sequential dependence of recurrent neural networks by processing words in parallel. The conceptual node is called "attention": it is a function that allows the model to evaluate words in a text differently, processing those most relevant to the model's understanding, returning an output that is the weighted sum of all individual values, which have been assigned a "weight" thanks to a specific function. In addition to the complete parallelization of the processes, Transformer manages to capture the relationships between related words even when they are placed in the sentence at a certain distance, consequently improving the performance of the model (Fig. 14)

Fig. 14 - Architecture of Transformer



Source: Vaswani, 2017

The Transformer architecture is the basis for the evolution of models such as GPT-3 and GPT-4. These are trained on huge text corpora, learning to predict the next word in a sentence: thus, they not only have the ability to handle the grammatical rules and vocabulary of a language, but also argumentative structures, semantic relations and even pragmatic elements of discourse. The training phase is functional for generating text autonomously from an input (a sentence, question or argument) that allows the model to generate subsequent text based on the knowledge gained in the training phase.

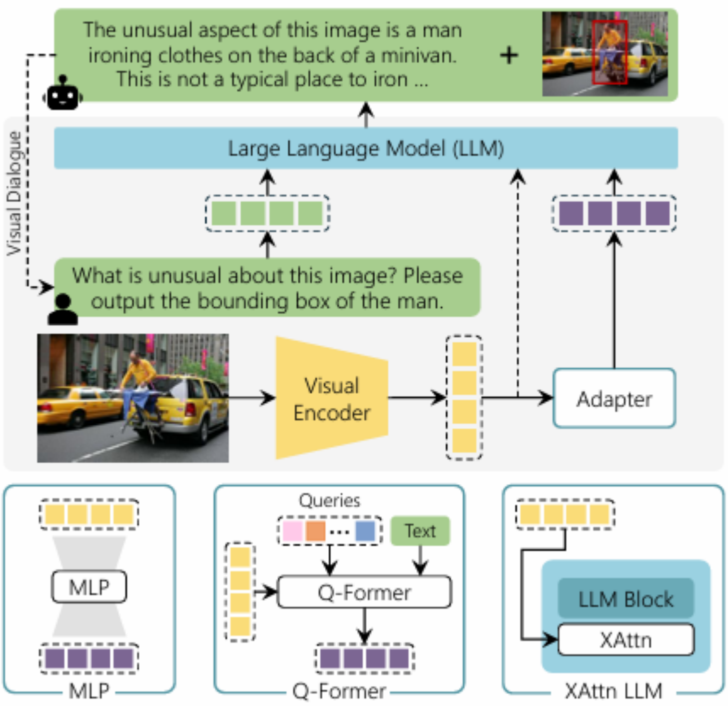
A crucial aspect of these models is their ability to generalize tasks for which they have not received specific training, using few or no examples. Specifically, these are models referred to as Few-shot learning (FSL) or Zero-shot learning (ZSL): it is a model's ability to perform a task without having seen any examples of it during training. Both models take advantage of the extensive pre-training of LLMs on diverse datasets, allowing them to perform new tasks without requiring the modification of interpretive parameters in order to arrive at a final result (fine-tuning). While the ZLS relies only on its pre-trained knowledge and ability to make predictions, the FLS has received pre-training with a limited number of examples from which it draws in order to make predictions. The FSL model generally performs better (Chen et al., 2022). These applications have opened up the possibility of using generative AI for a wide range of applications: virtual assistants, machine writing, machine translation, code generation, scientific and legal advice.

If LLMs have transformed our relationship with language, Multimodal Large Models (MLLMs) are transforming our relationship with reality in its sensory wholeness. These models combine text, images, video, and audio, fusing heterogeneous data into coherent, navigable representations. Multimodal models are designed to overcome the compartmentalization typical of specialized AI, adopting a flexible structure capable of interpreting complex contexts. For example, a model can read a text, compare it with an image, and generate a coherent response that takes into account information from both sources.

Technically, this is made possible by the use of multimodal encoders, shared attention layers, and training datasets with text-image or text-audio pairs. Multimodal models represent one of the most dynamic frontiers of AI research today. They are based on three fundamental components (Fig. 15):

- an LLM backbone that acts as an interface with the user;
- one (or more) visual encoder(s);
- one (or more) visual-to-language adaptor module(s).

Fig. 15 - Components and structure of an LLM



Source: Caffagni et al, 2024

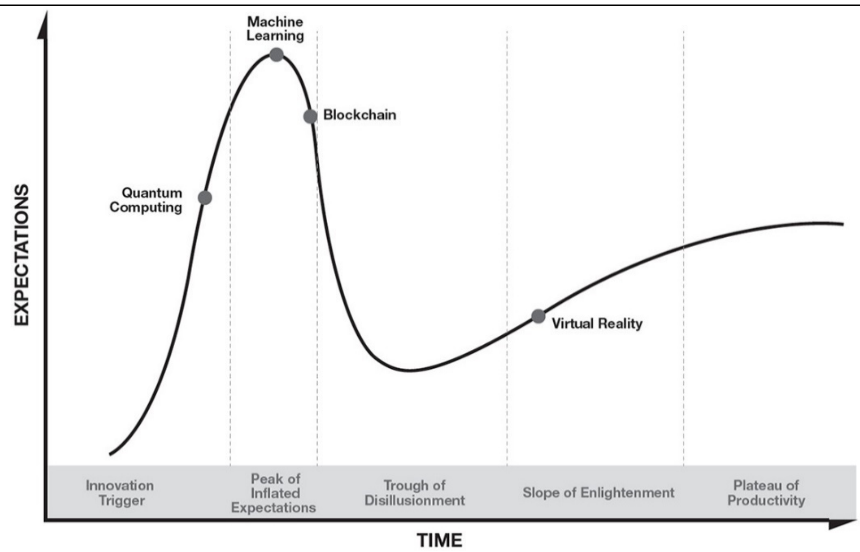
Part of the literature has pointed out that the evolution of architectures and models and the increasing rapprochement between perception and language delineate a vastly expanding and extremely dynamic field of operational research. With the centrality of text as the primary source of information, the generative capacity of MMLMs is developed in six key areas:

- **Text-to-Text (T2T**), which is fundamental to all tasks that have natural language as their output; the T2T mode is essential because it is the backbone of information management and for enabling access to it by conversational agents;
- **Text-to-Image (T2I**), capable of generating visual content and using it for image-based generative tasks;
- **Text-to-Music (T2M**), which represents one of the most complex models because of the richness of emotional and creative expressiveness;
- **Text-to-Video (T2V**), which combines temporal and visual information, generating dynamic outputs involving real-world physics, representing a complex model of reality;
- **Text-to-Human Movment (T2HM**), whose importance is growing because of its ability to be applied in robotics, animation, and virtual avatars, representing one of the most promising areas of development;
- **Text-to-3D (T2-3D)**, which links the ability to generate content with immersive environments, enabling exploration and interaction within immersive virtual spaces (Han et al, 2025)

This qualitative leap opens implications that go far beyond simple automatic content generation and carries far-reaching philosophical and practical implications. Indeed, a multimodal model is not limited to *understanding* a text, but can *see* an image, *hear* a voice, and most importantly, generate responses that synthesize these modalities and integrate them with each other. Generative artificial intelligence thus approaches an actualized form of cognition, reminiscent of human perception. These are aspects that lead many scholars to believe that generative AI has the ability to learn and adapt to contexts with cognitive mechanisms similar to those of human beings. The concept of AGI involves a cross-cutting capacity for problem solving, context adaptability, and continuous learning.Although some results are very promising, it should still be emphasized that current models lack intentionality, consciousness, sensitivity, and deep understanding of the world. Their *intelligence* remains highly dependent on training data and specific architectures designed by humans.

These considerations imply a reflection concerning the possibility that the domain of generative AI may be (if not today, then in the more or less near future) a *disruptive technology* or whether rather these systems, around which an enthusiasm and expectation has been created involving both scientific research and insofar as its application functions, should not instead be classified as a technological hype. The concept of technological hype refers to the enthusiasm and expectation that is created around a new product, service or trend that involves technological aspects and that is driven by a "promise" of the benefits that such an innovation can bring in terms of improvements in processes and their consequent impact on everyday life. Technological hype triggers a series of behaviors and attitudes that involve a very wide range of people (from specialists to researchers, and then implement a spillover to all individuals) and is often linked to marketing and its ability to generate expectations and promote needs in a specific sector. The cycle of a technology hype has found its theorization in the so-called "Gartner Cycle," which was defined by researchers at one of America's leading technology consulting firms. The next figure shows the trend of the cycle:

Fig. 16 - Gartner Cycle



Source: Smith, 2020

Every technology, upon its debut, experiences a period of enthusiastic reception by insiders and scholars, until it reaches a peak of expectations that are as high as, according to technology hype theorists, excessive in that they do not correlate with the real state of the art of the subject. A number of factors contribute to the attainment of this peak: marketing is supported by a strong focus by research technology, which sees the possibility of attracting funding and financing and consequently will be more oriented toward emphasizing positive results, while dissonant voices are altogether in the minority. The spike in expectations, especially if these are not or only partially realized, leads to an overall attitude of disillusionment with the real potential that the new technology shows. which, however, sooner or later are challenged in a long period of disillusionment with the real potential of the new technology. The impact with reality also generates a more rational realization of the new technology's ability to offer answers on a concrete level: this more reality-anchored attitude allows the transition from disillusionment to a growth in knowledge and awareness, which is consolidated with increasingly accurate (and perhaps more objective) studies and more effective knowledge of the subject matter. At this stage, the expectation curve rises again, albeit not as sharply as in the initial phase, and that preludes the achievement of an operational stability that represents the attainment of a maturity that preludes the effective diffusion of the benefits that the new technology brings (Fenn et al., 2008).

The classic double expectation curve described by Gartner's theory (an initial bell-shaped one, in which the apex is represented by the peak of expectations; an "S-shaped" one that preludes a normalization of the technology and is consistent with its "putting into production" in society) is discussed and controversial. The elaboration of this theory is mainly related to the development of Quantum computing, which in the representation of figure 15 is in fact in the ascent stage, at the apex of which machine learning is positioned, which today is instead, almost two decades later can safely be positioned in the productivity plateau stage. With respect to the applications of generative AI there remain (we will discuss this later in our paper) some difficulties related to ethical issues, particularly with respect to the management of the confidentiality of data conferred by users and their processing, and with respect to the ability of AI agents to manipulate people's behaviors, inducing cognitive and behavioral biases; the diffusion of AI Agents is already very wide, which suggests that the overcoming of current technological limits is destined sooner or later, thanks also to the advent of quantum supercomputers capable of speeding up computational capabilities.

The combination of LLM, multimodality, Reinforcement Learning (RL), Retrieval-Augmented Generation (RAG), and multi-step planning capabilities is increasingly bringing AI closer to an embryonic form of cognitive generation. The possibility that current Generative AI systems may develop a form of autonomous consciousness in a reasonably short period of time is a topic that meets the interest and concern of scientists. Although it is possible to envision a future in which AGI models can manipulate symbolic knowledge, learn causally, and develop the ability to reflect on their own internal processes (metacognition), it must be recognized that no current AI systems are conscious, but also that there are no obvious obstacles to building conscious AI systems.

Overall, we are probably at an intermediate stage: what we now call *generative AI* is a sophisticated sum of well-coordinated specialized modules, not yet a synthetic mind. But the trend is clear: evolution continues and requires critical monitoring and multilevel governance that balances technological progress and human values (Han et al., 2025).

2.3 What is an AI Agent: definition and characteristics

In the context of the evolution of artificial intelligence, one of the key concepts that has become increasingly relevant is that of "AI Agent." AI Agents are autonomous or semi-autonomous entities capable of sensing the environment, processing information, and acting to achieve specific goals. This essay explores the historical evolution of intelligent agents, defines the concept through authoritative sources, and analyzes its fundamental characteristics.

The notion of an intelligent agent has deep roots in the field of artificial intelligence, dating back to the early years of the discipline. As early as the 1950s, Alan Turing hypothesized the possibility of machines capable of intelligent behavior. However, it is only since the 1980s and 1990s that the term "agent" has acquired a more precise technical connotation, particularly with the emergence of distributed artificial intelligence.

In 1995, Michael Wooldridge and Nicholas Jennings defined agents along two main meanings of the term:

- a "weak" Agent, thus denoting a computer system that can act autonomously, without the need for human intervention and with some form of control over its internal processes, with the ability to interact with other agents (and of course with other humans) and to sense the environment (physical or virtual) by reacting to changes in it with timeliness and pro-activity;
- a "strong" Agent concept, which in addition to the previous functions also has the passibility of developing notions referable to human beings (emotions, empathy, intention and obligation).

Since the pioneering stage of AI Agent development, there has been some difficulty in defining AI Agents. The work of Wooldrige and Jennings that we have cited defines them precisely through their fundamental characteristics (autonomy, social skill, responsiveness and proactivity; to which are also added rationality, truthfulness and benevolence, i.e., the particular condition according to which agents do not have goals that conflict with what they are asked to do.

The first definition we propose dates back to the second half of the 1890s, when we were still talking about "autonomous agents" interacting with the real world. Frankline and Graesser proposed the following definition:

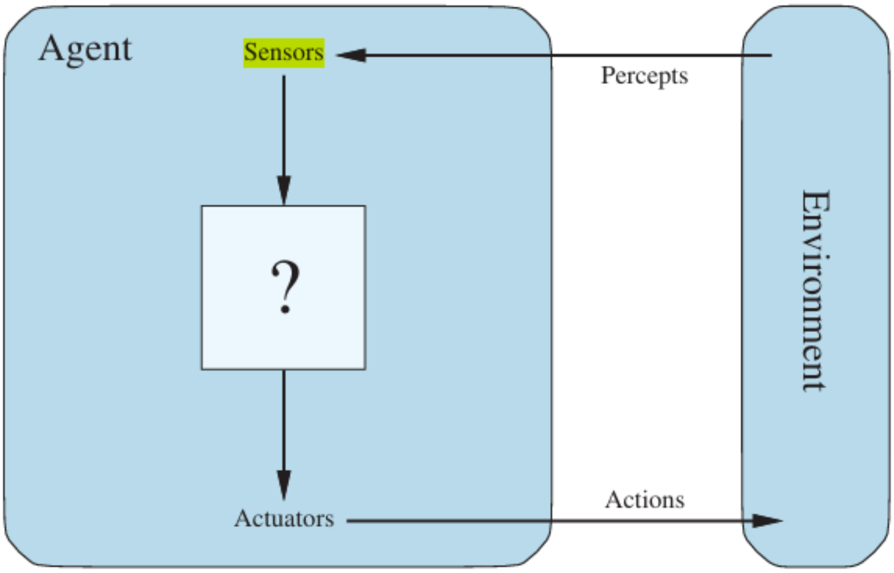
> An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future (Franklin & Graesser, 1997).

It is interesting to note two significant aspects of this definition. The first is that it is a synthesis proposed by the authors based on a number of previous definitions that represented well the state of the art at the time the article was proposed. Most of the definitions have the real word as a reference (Josè Carlos Brustoloni had also expressed himself in this sense, speaking of AI agents as "[...] systems capable of autonomous, purposeful action in real world." Brustoloni, 1991). The second significant aspect is that although dated, this definition has the merit of identifying the fundamental elements of modern AI agents. And in fact, more than two decades later, the attempt to define AI agents in a general key proposes a general definition of great simplicity:

> An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators (Russell & Norvig, 2020).

All of these definitions, as well as Wooldrige and Jennings' definition, have a common feature: although they are very different, they highlight the fundamental characteristics that define an AI Agent. At a general level, an agent is an entity that perceives its surroundings (through sensors, cameras, content received through files, access to banks from or even from human interaction, e.g., through keyboard or voice) and acts through actuators (e.g., electric motors, speech synthesizers, monitors and grippers in robotics; user interfaces and speech synthesis to determine speech output in software agents such as chatbots; motion commands, animations and sound effects of virtual agents in a vidogame). These actuators actually make actions possible, which in turn help generate interactions with the environment in which the AI Agent operates (Fig. 17).

Fig. 17 - Agent interactions with environments through sensors and actuators



Source: Russell & Norvig, 2020

This determines that AI Agents can also be classified and described based on these characteristics. The most significant are:

- **Autonomy.** This is the ability to operate without direct human intervention and is one of the main characteristics of agents. An autonomous agent is able to make decisions, adapt to the environment and modify its behavior over time. The capacity for autonomy is related to the ability to self-regulate behavior in response to environmental stimuli and goal variation and is a particularly important function in environments where human supervision of processes cannot be expected, such as in the case of chatbots for customer support (Sapkota et al., 2025);
- **Perception and sensors.** An AI agent is able to stand in its environment and perceive it through sensors (physical, in the case of robots; virtual, in the case of software). Perception has always been identified as a critical component for the development of AI agents (Nilsson, 1998), because it is the feature that enables the responsiveness and adaptability of the Agent; put another way, it is the perception of the environment that enables AI agents to implement actions that can achieve the goals
- **Knowledge representation.** The ability to interact with the environment is determined by the internal model of the AI Agent; this model dynamically updates to make interaction with the environment more effective. The representation of the environment in which it is located can be symbolic (when it is based on formal logic), probabilistic, and neural network-based connections;
- **Planning and Decision Making.** Agents must be able to select appropriate actions to achieve certain goals. This involves planning, optimization and sometimes, negotiation (in the case of multi-agent agents) skills. More advanced AI agent models place great emphasis on deliberative planning, helping to fuel the rationality of their behaviors (Russell & Norvig, 2020).
- **Learning capability** An intelligent agent is often equipped with machine learning mechanisms that allow it to improve its performance over time based on experience. This aspect allows for scalability of performance, which progresses based on

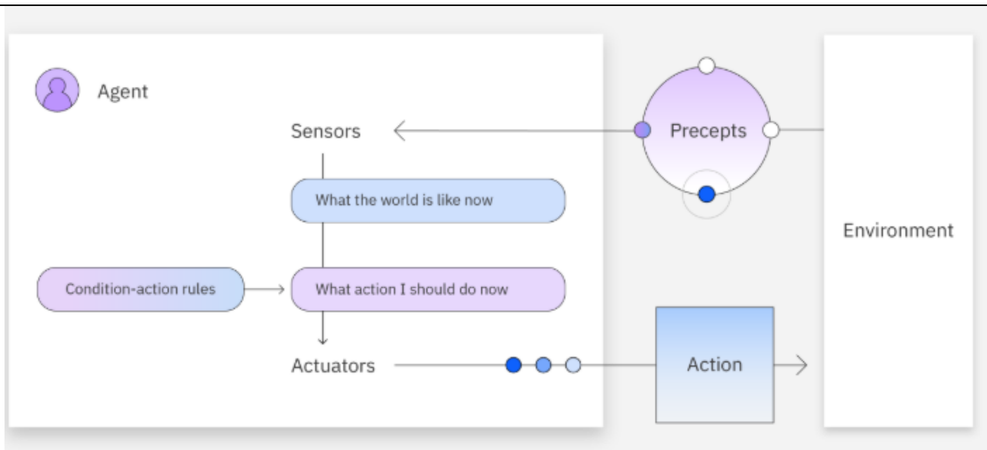
- **Social interaction.** In many contexts, agents must interact with other agents or humans. This implies the need for patterns of communication, cooperation and, in some cases, understanding of others' intentions. Social interaction is one of the fundamental characteristics of AI and extends to agents as well: in practice, this means that AI agents develop cognitive processes in a shared environment through processes of goal delegation. Based on social research and cognitive science findings, which see individual social actions and shared architectures as the focal point for modeling collective phenomena, AI Agents have also been provided with a design and architecture that can handle complex and intelligent interactions in dynamic environments (Sapkota et al., 2025).

2.4 Taxonomy of AI Agents.

We have seen how in the vast domain of artificial intelligence, the concept of agents (also defined as *autonomous agents* or *intelligent agents*, according to the perspectives that emerged at the early stages of AI theories) is declined in many ways, which have allowed the literature to divide the different categories of agents into six main classes:

1. **Simple Reflex Agents.** They base their behaviour solely on immediate perceptions, with no internal memory or representation of the world. They are the simplest form of agent and respond exclusively to immediate input, through a set of rules called "condition-action": classic examples of these systems are the thermostat (which turns heating on and off based on the interaction with the environment defined by the sensor that detects temperature) or a traffic light, which acts based on information from traffic sensors. They are effective agents in environments that record predictable behaviours that allow the application of well-defined rules; however, they show limitations in dynamic, high-complexity environments, partly because by not retaining memory of previous actions they do not develop any form of learning and long-term planning (fig. 18). Reactive agents show an additional problem: because their behaviour depends on predefined rules and do not store past actions, they can generate errors when situations occur in the environment that change the initial conditions (Stryker, 2025)

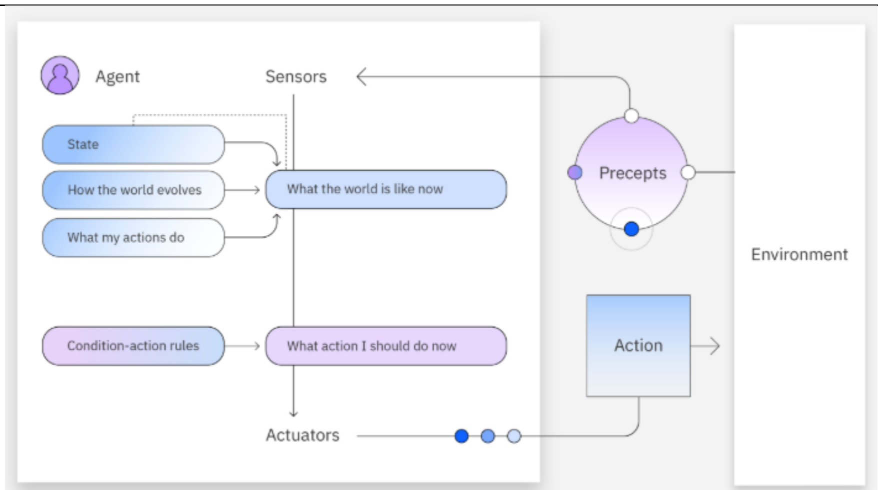
Fig. 18 - Reflected agent model



Source: Stryker, 2025

2. **Deliberative agents.** This class of agents achieve the goals for which they are predisposed by having as an internal reference a symbolic model of the world in which they are immersed; this allows them to interpret perceptions coming from outside even under conditions of partial observability, taking actions through symbolic reasoning (Russell & Norvig, 2020). Deliberative agents can also act on the basis of predefined rules, analogous to what happens with respect to reflexive agents; but incorporating a model of the reality with which they interact and track past states to respond to stimuli that come from environments that can only be partially observed.

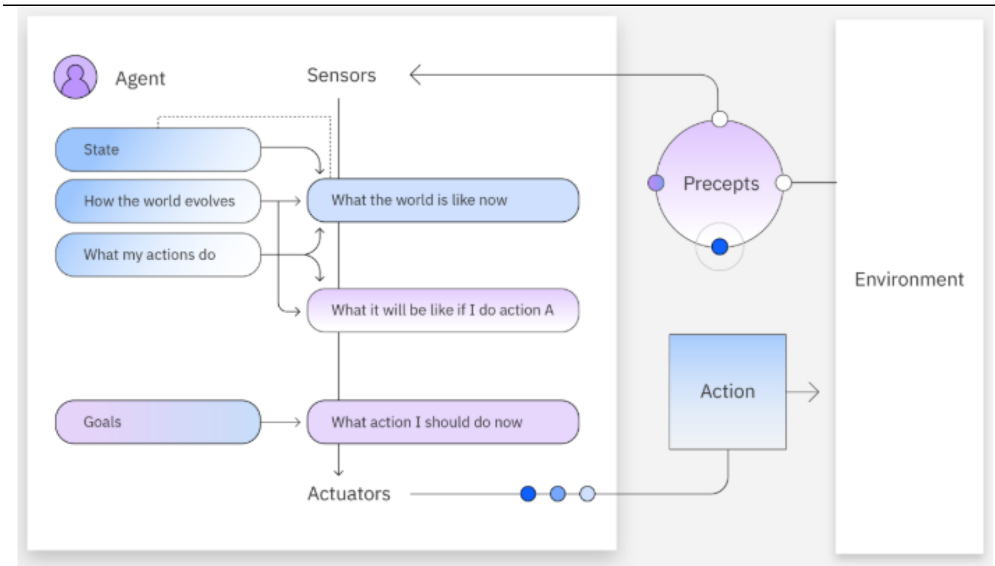
Fig. 19 - Deliberative agent model



Source: Stryker, 2025

3. **Goal-oriented agents.** They act to achieve specific goals, which can be expressed in logical form or as functions to be optimized. The operation of these agents depends on the application of classical or probabilistic planning algorithms. Compared to the previous classes, these agents pursue multiple goals simultaneously, depending on a utility defined in the algorithms. They are more effective than agents of the previous classes in terms of their ability to operate in dynamic scenarios, where constraints and rules change over time. In addition to maintaining memory of past states, a goal-oriented agent is able to evaluate future states and assess their impact on the achievement of goals. Their use is common in robotics, self-driving vehicles, and simulation systems: environments that combine clear goals with the ability to adapt with respect to decisions to be made in real time. However, they have the limitation of depending on predefined strategies and decision trees, which guide the agent's behaviour (Fig. 20)

Fig. 20 - Goal-oriented agent model

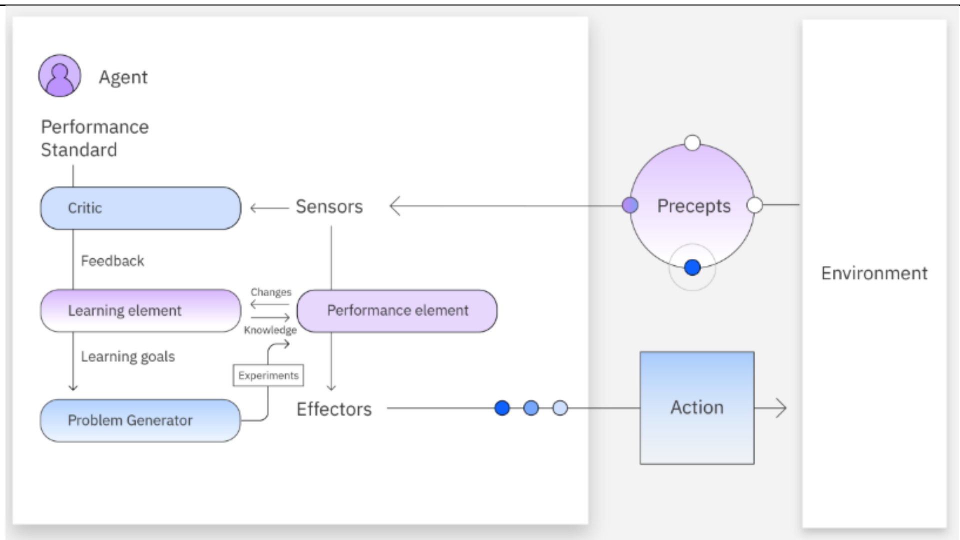


Source: Stryker, 2025

4. **Learning agents.** They improve their performance over time using machine learning techniques, including supervised, unsupervised and reinforcement learning. This type of agents is based on their ability to update their decisions according to feedback from the environment: this determines the ability to

improve decision-making capabilities particularly in environments characterized by uncertainty and dynamic events. Their architecture is more complex (Fig. 20): in addition to the components present in the other classes of agents, they make use of a learning element (which improves the agent's knowledge based on feedback and experience), a so-called "critical" element, which has the task of evaluating the agent's actions and providing feedback with respect to the decisions made, and a problem generator, which has the function of suggesting potential scenarios that will cause the agent to discover new strategies and improve its knowledge. These features imply flexibility and the ability to handle complex and constantly changing environments. This is a class of agents particularly used in chatbots and social media, due to the potential in natural language processing that analyzes user behavior to provide and suggestions on relevant and optimized content (Stryker, 2025).

Fig. 21 - Learning agent model

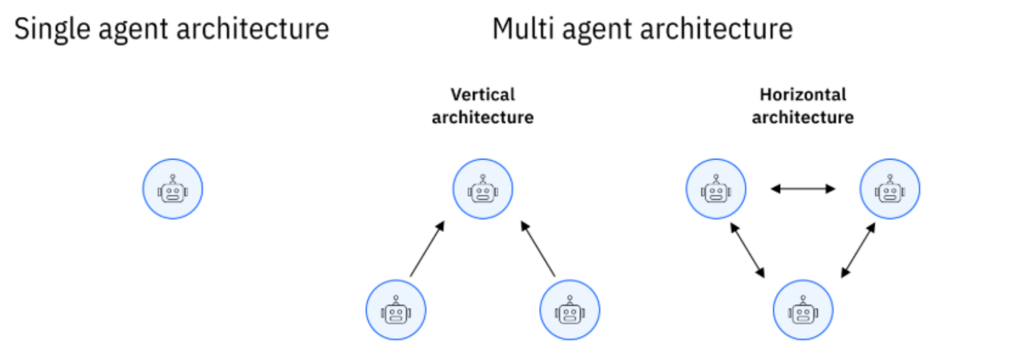


Source: Stryker, 2025

5. **Multi-agent systems**. This class of agents was created with the aim of effectively responding to the challenges of the complexity that characterizes not only productive reality, but also AI systems. Indeed, they decompose complex problems into a series of simpler tasks, which are therefore easier to manage. The decomposition of problems into a series of tasks leads to greater possibilities of

responding to the stresses of the real environment: agents can in fact be dedicated to higher tasks (setting general goals and strategies) and to more specific tasks (Sapkota et al., 2025). The relationship between different agents can be hierarchical or linear (Fig. 22):

Fig. 22 - Multi-agent systems



Source: Stryker, 2025

This albeit cursory classification allows us to grasp one of the potentially most important evolutionary directions: the transition from AI Agents to an agentic AI. Indeed, the need to manage complexity implies the expansion of advanced components (Specialized Agents, Advanced Reasoning and Planning, Persistent Memory, and Orchestration). These components cooperate within a shared ecosystem, in which task decomposition and coordination of agent functions effectively create a new collaborative paradigm marking the transition to a distributed and adaptive agentic AI intelligence (Sapkota et al., 2025).

3. AI AGENTS: ARCHITECTURE, OPERATION, AND BUSINESS INTEGRATION

3.1 Structure and components of an AI agent: planner, tools, memory, feedback loop

An AI agent, as an autonomous entity capable of perceiving its environment and acting on it to achieve predefined goals, is inherently endowed with a complex architecture that enables it to function intelligently and adaptively: the effectiveness and robustness of this architecture is fundamental to enabling the AI agent to perceive the environment, process information, make decisions, and act autonomously or semi-autonomously. Unlike passive artificial intelligence systems that process information without direct interaction, AI agents are autonomous systems capable of interacting with their environment, making decisions, and taking actions to achieve specific goals. For this reason, an AI agent is composed of several interconnected components that work together in synergy to achieve the goals to be achieved. It is this architectural complexity that integrates the modular components that support perception, cognition, and goal implementation, allowing the agent to interpret the context, reason, and plan its actions in dynamic environments (Bousetouane, 2025).

The planning module is one of the fundamental pillars of this architecture and is effectively the 'brain' of the AI agent. Its main function is to formulate strategies and sequences of actions to navigate and operate effectively within its operational domain. The Planner is responsible for translating high-level objectives into detailed operational plans, often using search algorithms and symbolic reasoning techniques to explore the space of possibilities and select the most efficient path: for example, it can break down complex tasks into simpler sub-tasks and define a logical sequence to achieve the objective (Latapie et al., 2022). This process involves evaluating different alternatives, predicting their outcomes, and adapting plans based on changing environmental conditions or new information, and is critical to enabling the AI Agent to operate in a multi-step perspective in highly dynamic environments (Sapkota et al., 2025). Planning can be pursued through different methodologies:

- **Chain-of-Thought (CoT) Reasoning**: this methodology, inspired by human reasoning, allows the AI Agent to break down complex problems into intermediate steps, formulating a logical chain of thought that leads to the final solution, increasing the reliability of decisions and the transparency of the decision-making process. With regard to LLM-based AI Agents, the literature has highlighted three main planning strategies aimed at improving the adaptability of the agents themselves based on a more or less deep interaction with the user:

1) The "Planning-Only" approach directly generates activity plans based solely on the user's initial prompts, without actively engaging in further clarification; this is the approach adopted by most existing DR agents;
2) The "Intent-to-Planning" strategy seeks to proactively explore the user's intent before proceeding with planning: to do this, it asks the user targeted questions, subsequently generating customized sequences of activities based on the user's clarified inputs; this method is used by OpenAI DR
3) The "Unified Intent-Planning" approach combines the two previous methods: after generating a preliminary plan based on the initial request, it interactively involves the user to confirm or reassess the proposed plan. Gemini DR is the best-known example of this strategy, whose strength lies in user-guided refinement (Huang et al., 2025).

- **Self-reflection**: the ability to self-reflect allows the agent to critically evaluate its own plans and actions, identifying potential errors or inefficiencies and making the necessary corrections to optimize performance. The capacity for internal reflection can be conceptualized as an iterative process of self-assessment and self-correction, which allows the agent to refine its strategies and improve the effectiveness of its actions over time. from a technical point of view, it manifests itself through the analysis of deviations between expected and actual results, triggering mechanisms for the revision and optimization of internal models or heuristics used for planning (Yüksel & Sawaf, 2024).
- **Adaptive learning**: this approach allows the agent to refine its strategies and improve the effectiveness of its actions over time, especially in dynamic and not fully predictable environments: past experiences are generalized and used to improve performance by integrating adaptive learning with reinforcement mechanisms or neural networks (Chen et al., 2024). An example of such an architecture is the Talker-Reasoner framework, where a "Talker" agent handles quick and intuitive conversational responses, while a "Reasoner" agent handles

multi-step reasoning, planning, and execution of actions in the real world: the distinction between the two components offers the possibility of optimising the responsiveness of the "talker" component with the strategic reasoning ability of the "reasoner" component (Christakopoulou et al., 2024).

Interaction with the outside world is ensured by tools, through which the AI Agent is able to interact with and manipulate its operating environment based on external inputs. These tools can range from simple programming interfaces to much more complex applications, extending the agent's operational capabilities far beyond its intrinsic skills. In fact, tools represent a functional extension of the agent, allowing it to perform specific operations or access external resources necessary to achieve predefined objectives (Garg et al., 2024). Tools may include:

- **APIs (Application Programming Interfaces):** To connect to and manipulate external software such as CRM (Customer Relationship Management), ERP (Enterprise Resource Planning), calendar systems, email, messaging platforms (Slack, Teams), and databases. These programming interfaces, whether based on RESTful, SOAP, or GraphQL protocols, allow the AI agent to interact programmatically with web services, databases, and other software applications (Gosmar et al., 2024).
- **Specific functions:** Predefined capabilities that the agent can invoke for specific tasks, such as generating reports, analyzing data, or sending notifications. These functions can be internal to the agent or call on external services specialized for tasks such as natural language processing, computer vision, or predictive analytics.
- **Sensors:** In the case of robotic or physical agents, sensors are the input mechanisms that collect data from the environment (AWS, n.d.). They convert physical stimuli, such as light, sound, temperature, or pressure, into electrical or digital signals that the agent can process to perceive the state of the surrounding world. Through the use of these sensors, the AI Agent is able to build an up-to-date internal representation of its environment, which is essential for informed decisions and appropriate actions.

An effective AI Agent is able to select and use the most appropriate tool based on the context and objective, demonstrating practical intelligence in interacting with the surrounding environment to achieve its goals (Jovanovic et al., 2025).

The ability to store and retrieve relevant information produced by past interactions (e.g., user preferences or the history of past activities and choices) is ensured by *memory*. Memory provides the context for making decisions in the present and future, ensuring the agent's consistency and learning over time and allowing it to build cumulative knowledge that enables it to refine its strategies in the future (Cichocki et al., 2021). Generally, AI agents have a two-level memory:

- **Working Memory (Short-Term/Working Memory):** this is a short-term memory specific to the current session or active task. It contains temporary information such as the current conversation, the user query, or the current status of the task. It is a volatile memory and is typically deleted after the task is completed.
- **Persistent Memory (Long-Term Memory):** this memory allows the agent to recall historical information from multiple past sessions and is developed through external vector databases or other forms of long-term storage; its role is crucial in learning, adaptation, and improving agent performance over time, as it allows the agent to "remember" important interactions, user preferences, or changes in the environment.

This distinction between short-term and long-term memory is fundamental to the flexibility and effectiveness of AI agents, allowing them to manage both immediate needs and the accumulation of long-term knowledge for complex tasks (Gupta et al., 2025).

The *feedback loop* is an additional component of AI agents that is necessary to reinforce the process of learning from the results of their actions and adapting their behavior accordingly. The feedback loop is fundamental to the evolution and continuous improvement of the agent's performance: it allows it to evaluate the effectiveness of its decisions and modify future strategies based on past results, evaluated through responses obtained from the environment (from interactions with the user or from the effects of the results obtained) (Wayne et al., 2018).

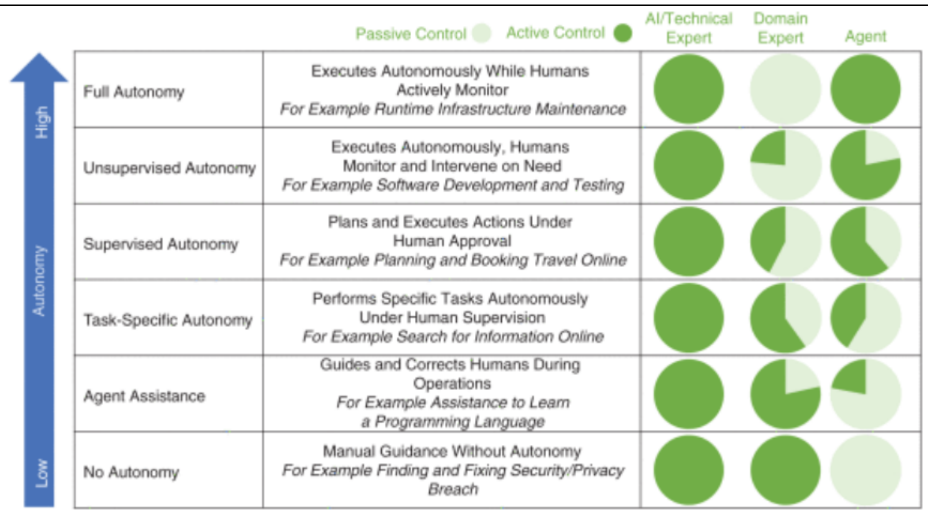
The feedback loop process typically includes:

- **Perception/Input:** The agent receives an input or "trigger" from the environment, which can be a new request, a message, or an API call.
- **Execution:** The agent performs the planned actions using the available tools.
- **Results monitoring:** The agent monitors the success or failure of the actions taken.
- **Analysis and Learning:** The agent's learning component analyzes past actions and their results to modify its internal knowledge base and improve future results. This can be done through supervised, unsupervised, or reinforcement learning techniques.
- **Adaptation:** Based on feedback, the agent can refine its planning, tool selection, or memory management to optimize future actions. This continuous cycle of evaluation and refinement is intrinsic to the architecture of AI-based agents, allowing them to progressively improve decision accuracy and operational efficiency (Anderson et al., 2021).

3.2 Types of AI agents: autonomous, collaborative, hybrid

AI agents can be divided into different categories based on their level of autonomy and ability to interact with other agents and other systems (including interaction with humans and, in particular, with the consumer to whom the suggestions are directed). AI agents are created and trained with a predefined level of autonomy, which may involve entrusting a more or less significant part of the control and supervision to human domain experts. This 'share' effectively results in a loss of autonomy for the AI agent: if we consider the combination of autonomous control carried out by the Agent and that carried out by experts, we obtain a typology of AI Agents, shown in Figure 23, which distinguishes between autonomous agents, which operate without human supervision for long periods and make independent decisions, and collaborative agents, which interact and cooperate with other agents, whether human or artificial, to achieve common goals.

Fig. 23 – Levels of autonomy of AI agents

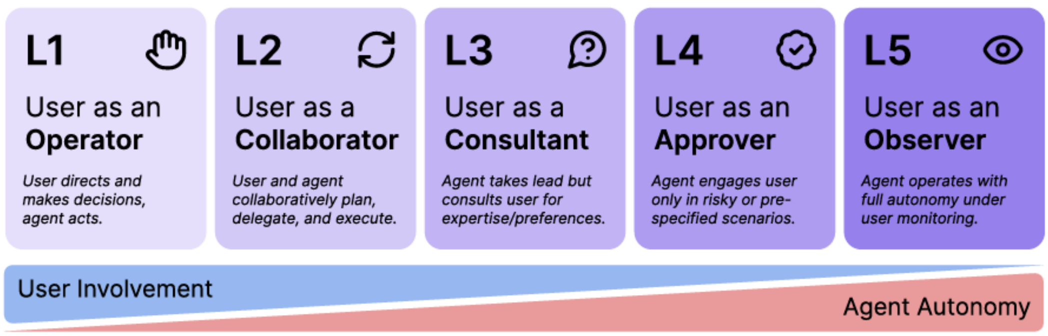


Source: Jovanovic et al., 2025

As can be seen, except in extreme cases (complete autonomy or no autonomy), agents combine the characteristics of both types, operating autonomously in certain contexts but integrating collaboration for complex tasks or to optimize performance (Jovanovic et al., 2025).

A further subdivision of AI agents based on their level of autonomy and user involvement has been described on the basis of five distinct levels which, as in the previous categorization, imply a different aggregation between user and agent (Feng et al., 2025). Figure 24 illustrates this taxonomy:

Fig. 24 – Taxonomy of AI agent autonomy levels



Source: Feng et al., 2025

In this case, the starting point of the taxonomy is the user, and the scalability of their involvement ranges from total involvement in the decision-making process, which remains entirely at the user's disposal and entrusts the AI Agent with the task of performing the assigned tasks, to a mere observer role, which entrusts all decision-making functions to the Agent.

The need to distinguish AI Agents according to their type is not only descriptive: it is in fact functional in order to choose the agent best suited to the specific requirements of the task and to the needs of the company, and to manage its impact both in production processes and in the equally complex world of customer and stakeholder relations. The literature distinguishes between autonomous agents, collaborative agents, and hybrid agents.

Autonomous agents are intelligent systems designed to perceive the environment, reason, and act independently in order to achieve predefined objectives without the need for constant human intervention (Ren et al., 2025). Their autonomy derives from their ability to integrate modules such as *planners, tools, memory*, and *feedback loops* to perform complex tasks from start to finish (Xi et al., 2023).

The distinctive characteristics of autonomous agents include:

- **Decision-making independence:** autonomous agents are able to choose the most appropriate actions based on their internal state and perceptions derived from interaction with the agent, recorded through feedback that is entrusted to them; the need for human supervision is minimized, although not entirely eliminated, as it is still essential in setting the guidelines or information that constitute the initial input of the process (Ren et al., 2025);
- **Adaptability and Proactivity:** autonomous agents are able to adapt to changes in the environment through learning and feedback mechanisms, and are capable of initiating actions to achieve their assigned objectives; they are therefore not limited to simply responding to stimuli from the user, but adopt autonomous strategies to achieve the expected results (Xi et al., 2023);

According to the taxonomy of user-agent interaction proposed in Figure 23, autonomous agents operate at autonomy levels more like L4 (*Approver*) or L5 (*Observer*): at level L4, the agent plans and executes most tasks, consulting the user only for approval of consequential actions or in case of failure/high-risk states, while at level L5, the agent plans and executes all tasks, with the user acting primarily as a monitor and auditor, having only an "emergency switch" at their disposal with no means for direct involvement otherwise (Feng et al., 2025).

The ability of autonomous agents to operate without direct human intervention makes them crucial for the automation of complex processes and for resource optimization in highly specialized sectors: typical examples of autonomous agents are self-driving vehicles, industrial robots that perform complex assembly tasks, or supply chain management systems that optimize logistics flows in real time. In a business context, an autonomous agent could manage the entire cycle of a marketing process, from data analysis to personalization and automatic campaign execution, with minimal human supervision. The implementation of autonomous agents raises critical questions regarding the calibration of appropriate levels of autonomy, which must be carefully balanced to mitigate risks and ensure effective governance: these systems, in fact, present significant challenges in terms of safety (Talukdar, 2025).

Collaborative agents, on the other hand, are designed to operate in close synergy with humans or other AI agents, with whom they cooperate with the aim of achieving a common goal. Their effectiveness lies in their ability to interact, communicate, and coordinate their actions with other entities, making the most of both the computational capabilities of AI and human intuition and decision-making abilities.

The key features of collaborative agents include:

- **Interaction and communication:** collaborative agent-based systems use interfaces and protocols to develop effective, simple, and as rewarding as possible two-way communication with users. These platforms are used to exchange information, plans, and statuses necessary to achieve predefined objectives.
- **Information sharing:** collaborative agents exchange relevant data and knowledge with their partners, both human and artificial, to ensure optimal coordination of activities (Xi et al., 2023).

Collaborative agents typically fall within levels L2 (Collaborator) and L3 (Consultant) of the taxonomy in Figure 23. At level L2, the user and agent plan, delegate, and execute tasks jointly, with the agent proactively suggesting actions, although the user retains full control of the process. In level L3, the agent suggests actions that must be approved by the user before being implemented; the agent performs most of the tasks, but in doing so adopts a strategy that involves constant collaboration with the user, who contributes to the achievement of the result with their expertise or by using specific preferences (Feng et al., 2025).

Collaborative agents are inherently designed to address problems that would be very difficult for a single isolated agent to solve: they thus fall into the category of multi-user systems and promote knowledge sharing and cooperation between intelligent entities for the emergence of consistent and innovative solutions (Cichocki et al., 2021). Their architecture reflects these approaches: it is based on frameworks that facilitate communication, negotiation, and joint planning, enabling the integration of different skills for solving complex problems in distributed contexts. Typical examples of collaborative agents are advanced chatbots that assist customer service, AI agents that support designers in generating creative ideas, and virtual assistants in companies that collaborate with managers for project planning and optimization. This highlights one of the most significant features of this category of agents, namely their ability to amplify human capabilities, allowing individuals to focus on higher value-added activities and leaving more automatic operations to AI agents.

Hybrid agents represent a synthesis of the characteristics of autonomous and collaborative agents: they provide varying degrees of independence and interaction depending on the context and the specific task requirements. This type of AI agent is particularly relevant in more complex environments, where some activities can be fully automated, while others require closer human supervision or interaction to ensure accuracy, compliance with service or product standards, or to effectively manage exceptions (Stryker, 2025).

The main characteristics of hybrid agents are:

- **Operational Flexibility:** hybrid systems are able to dynamically switch from a nearly fully autonomous mode of operation to a collaborative mode; the switch is made by assessing the complexity of the task, the risks associated with it, or the need for interaction with a human user or another system (Ren et al., 2025);

- **Modularity and Scalability:** hybrid agents are designed with operational flexibility in mind and therefore feature a modular architecture that allows the degree of autonomy and interaction to be configured according to specific requirements, thus offering greater scalability;
- **Work Optimization:** the hybrid approach aims to optimize overall efficiency: it allows repetitive and well-defined tasks to be delegated to the AI agent, while strategic decisions, exception management, or the interpretation of ambiguous or previously unknown data are reserved for human resources;

A practical example of a hybrid agent could be a sales management system that autonomously identifies qualified leads and generates initial proposals, but requires the intervention of a human salesperson for final negotiation and contract closure. Similarly, a quality control agent in manufacturing could autonomously monitor processes and report anomalies, but delegate the resolution of complex or unexpected problems to a human technician (Ren et al., 2025). The choice of a hybrid agent is linked to the need to balance efficiency and reliability of AI agents with the ability to manage the complexity and novelty that is typical of more complex socio-economic ecosystems.

3.3 Integration into business flows: CRM, ERP, decision support

Artificial intelligence (AI) has taken on the role of a transformative force in the modern business landscape, driving organizations toward deeper integration of AI technologies into their operational and strategic processes. Contemporary economic ecosystems are increasingly based on the analysis of large amounts of data and process automation, machine learning, and natural language processing can be fundamental supports to help businesses maintain competitiveness in the medium to long term. AI agents have established themselves as essential tools: in increasingly complex economic contexts, they can leverage their tool and memory capabilities to interact with the APIs and databases of existing business systems, such as Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP), and in general with all decision support platforms in use (Hlatshwayo, 2023).

A first strategic approach involves integrating AI agents into CRM systems. Although often simplified as management software, CRM is actually much more complex: the main definitions found in the literature describe it as an "approach" or "process" aimed at optimizing a company's profitability through the complex system of relationships it maintains with consumers and customers. Among the many possible definitions, we suggest the one proposed by Francis Arthur Buttle and Stan Maklan:

> "CRM is an integrated approach to identifying, acquiring, and retaining customers. By enabling organizations to manage and coordinate customer interactions across multiple channels, departments, lines of business, and geographies, CRM helps organizations maximize the value of every customer interaction and drive superior corporate performance" (Buttle and Maklan, 2015).

The definition we have proposed has the merit of introducing some fundamental elements of CRM: the management support function ("to manage and coordinate the "), the multi-channel nature of communication with actual and potential customers, and above all its role as an element of synthesis and integration: to be effective, this tool requires strong integration between different operational and strategic elements and the ability to combine relationship marketing approaches with the potential represented by corporate information systems; the process overseen by CRM is also based on the ability to analyze ever-increasing amounts of data in order to define the consumer target with ever-greater precision, also using information on the behavior and preferences of individual customers. The ability to manage large amounts of data has revolutionized customer interaction management at all stages of the process, and through intelligent automation and large-scale personalization, AI agents help improve the overall customer experience and business efficiency in both the sales and after-sales and support phases.

The scope of application of AI agents integrated with CRM allows for:

- **enhance customer service:** AI agents, in the form of chatbots and voicebots, can independently handle a wide range of customer service requests without being limited by time or day of the week, as they can operate 24 hours a day, 7 days a week; thanks to their potential, they can provide immediate answers, resolving common and standardized problems and leaving only the most complex cases to human operators. Collaborative AI agents can act at the "Collaborator" or "Consultant" levels (Feng et al., 2025), pre-processing customer requests and providing human operators with contextual summaries and response suggestions, thereby improving problem resolution times and service quality (Hlatshwayo, 2023). This not only increases efficiency, but also frees staff from routine and repetitive tasks, allowing them to be deployed in areas where greater empathy and the ability to handle complexity are required.

- **Automate and optimize sales:** AI agents can analyze large sets of customer data to identify the most promising leads, classify them according to predefined criteria, and automate the sending of personalized communications and offers. They can also monitor customer behavior in real time and suggest proactive actions to the sales team: follow-ups and cross-selling based on predictive analysis of purchasing patterns and preferences contribute to significant gains in operational efficiency (Basiru, 2023).
- **Build customer loyalty through personalized support.** Thanks to the long-term memory of learning algorithms, AI agents can build and maintain extremely detailed and up-to-date customer profiles. This in-depth knowledge enables the creation of hyper-personalized marketing campaigns and tailored offers, increasing customer engagement and strengthening brand loyalty (Hlatshwayo, 2023).

Like CRM, ERP also has a systemic character. It can in fact be considered

> "[...] a cross-functional enterprise system that integrates and automates organizational core business processes for achieving efficiency and effectiveness" (Bokhari, 2019).

From a practical point of view, an ERP can be described as a set of integrated software modules connected to a central database that allows the integration of information contained in each of the modules: the introduction of these tools into the company organization has progressed over time in relation to the spread and refinement of information systems based on technological resources. The biggest challenge in this area is the integration of information from different databases: ERP systems can guarantee control of the many processes implemented in the organization, but they require a reengineering effort to enable real-time sharing and analysis (Bokhari, 2019). The strong demand for the management of large amounts of data from a heterogeneous galaxy of systems is the conceptual basis that has driven the integration of AI agents into ERP systems: artificial intelligence has proven to be fundamental in optimizing internal operational processes, improving efficiency and reducing costs, offering a unified and accurate view of company resources (Maldonado-Canca et al., 2024). AI agents can harmonize different business processes, automating complex workflows that span different components of the company structure.

The main applications of AI agents in ERP are:

- **Intelligent Supply Chain Management:** AI agents can offer strategic support at all stages of the production process: for example, they can help monitor inventory levels, forecast demand more accurately, and automate replenishment orders, proactively identifying potential disruptions or inefficiencies in the supply chain (Hlatshwayo, 2023). They can also optimize logistics by tracking shipments and managing exceptions in real time, ensuring leaner and more resilient operational flows.
- **Financial Process Automation:** Agents can automate repetitive, high-volume tasks such as invoice reconciliation, expense management, payment processing, and anomaly or fraud detection. They also support the preparation of financial reports by aggregating and analyzing data from different sources within the ERP system, increasing accuracy and reducing human error (Basiru, 2023).
- **Human Resources (HR) Optimization:** In the HR context, AI agents assist in processes such as candidate screening, employee request management (e.g., vacation scheduling, payroll information management, etc.), and providing automated responses to frequently asked questions. The use of AI agents in more repetitive tasks frees HR staff from routine administrative tasks, allowing them to focus on strategic initiatives such as talent development and corporate culture.

In addition to operational automation, AI agents are also effective decision-making tools: they can provide insights based on complex data analysis and facilitate predictive and prescriptive analysis. They have the ability to process large datasets, identify hidden patterns within them, and simulate future scenarios: functions that are indispensable for top management when making decisions at both the strategic and tactical levels (Hlatshwayo, 2023).

The main key aspects of the decision support offered by AI agents can be identified in the following functions:

- **Advanced data analysis and insight generation:** AI agents are capable of collecting and aggregating data from disparate sources (CRM, ERP, external markets, social media) and, based on this, identifying correlations that might otherwise be missed, presenting intelligent reports and customized dashboards to management and stakeholders. Transforming raw data into functional and analytical insights improves the accuracy of decision-making (Basiru, 2023).
- **Predictive and prescriptive modeling:** using machine learning algorithms, AI agents can predict market trends, sales performance, operational risks, or the success of new initiatives. Their prescriptive capabilities extend to suggesting optimal actions for achieving specific objectives, such as price optimization, investment strategies, or production planning;
- **Scenario simulation and risk assessment:** AI agents can simulate the impact on business performance of decisions made by top management or of economic, regulatory, and market changes that alter the economic and social ecosystem in which the company operates. This analytical capability allows decision-makers to assess potential outcomes and mitigate risks by developing contingency plans before implementing new strategies (Basiru, 2023).

The effectiveness of integrating AI agents into these systems is strongly influenced not only by their technical ability to connect to platforms, but also by the clarity of business objectives, the quality and accessibility of data, and the organization's ability to manage the challenges of AI anxiety, trust, and workforce adaptation to new processes (Maldonado-Canca et al., 2024; Basiru, 2023).

It is interesting to note that the integration of AI agents into production processes is no longer an exclusive privilege of large companies, but represents an opportunity that is also within the reach of SMEs. This is made possible by two fundamental characteristics that have reduced the barriers to entry for AI solutions, even for small businesses.

The first element is accessibility. Until recently, using artificial intelligence systems required significant initial investments in terms of both architecture and specialist skills. the spread of cloud-based platforms and the availability of solutions that offer software as a service rather than as a commodity have enabled SMEs to access pre-trained AI tools that can be easily configured via APIs, allowing them to integrate these tools into existing workflows without the need for specialists. This ease of use has become a strategic asset for the spread of AI in SMEs, which often face budget and resource constraints and tend to take a more cautious approach than large companies to the potential uses of AI (Maldonado-Canca et al., 2024). Low-code/no-code solutions and plug-and-play AI agents have also enabled non-technical staff to implement intelligent automation, democratizing access to advanced data analysis and process automation capabilities that were previously prohibitive. The ability of agents to operate at different "Levels of Autonomy" (Feng et al., 2025) and the possibility of feeding human supervision or collaboration processes has made adoption less risky and more controllable for SMEs, supporting their tendency to approach the benefits of automation in an incremental manner.

This last consideration highlights the second key factor that has driven the adoption of AI solutions in SMEs. The scalability of AI agents is a key factor in sustainable growth: modern agent architectures, which are often modular and cloud-based, allow AI solutions to expand or contract according to business needs, both in terms of data volume and operational complexity. This means that an SME can start with the automation of a single process, such as customer service (e.g., adopting a chatbot for FAQs), and then extend the use of agents to other areas, such as lead management in CRM or inventory optimization in ERP (Hlatshwayo, 2023). Scalability is also reflected in the ability of agents to learn and continuously improve as data increases, enabling progressive optimization of business processes and decision- s (Basiru, 2023). This operational flexibility ensures that the investment in AI continues to generate value as the company grows, transforming AI solutions from a fixed cost into a variable cost that can be adapted to specific needs. However, some fundamental issues remain unavoidable and must also be considered by SMEs: challenges related to data quality and concerns about security and ethical principles in the use of AI can impact the scalability and long-term success of AI adoption by SMEs, which are still required to plan and invest in training for their internal staff (Basiru, 2023).

3.4 Ethical implications, security, and governance

The rise and pervasive integration of AI agents in business processes and everyday life raise questions that go beyond mere technological efficiency, making it necessary to carefully analyze the ethical implications of their use and address the challenges associated with the governance of AI agents on the one hand, and security on the other. Ethics, security, and governance are the three pillars on which the ability to integrate the enormous development potential of AI (a potential that is still not fully explored and rapidly growing) with the necessary protection of individual rights and the confidentiality of information collected about them is based. The debate on these issues is very heated and rich in food for thought: it has the task of promoting and managing the healthy development of AI and of support tools for production and marketing processes that are truly oriented towards the collective well-being (Zang et al., 2024).

The ethical implications of AI agents are complex and multifactorial. Artificial intelligence is based on algorithms: these are not 'neutral' because they are designed to maximize certain results chosen by the developer, which in itself involves prioritizing some interests over others. Although algorithms can be developed according to defined and accepted regulatory parameters, it is not easy to determine the ethical impact they may have: on the one hand, the ability to learn autonomously in unsupervised processes carries the risk that they may adopt strategies that inadvertently violate fundamental rights, for example by inadvertently profiling the population based on perceived ethnicity in order to offer more appropriate commercial recommendations; on the other hand, they are able to influence individuals' decision-making processes, even beyond the actual intentions of developers, thanks to their ability to modify operating parameters and decision-making rules. The use of learning algorithms therefore impacts a series of fundamental principles on which civil coexistence is based: fairness, transparency, accountability, and social impact may be profoundly and problematically affected and represent the main ethical challenges that the development of AI poses to developers and users. The inherent ability of AI agents to make autonomous decisions and operate in complex environments makes managing these challenges a priority. A significant risk is the induction of cognitive and behavioral biases in consumers: AI agents, learning from large data sets, may inadvertently reflect and even amplify existing social prejudices, leading to discriminatory outcomes in sensitive areas such as employment dynamics, access to credit, or opinions about justice and security (Mittelstadt et al., 2016). Ensuring the quality and fairness of training data and developing effective mechanisms for identifying and mitigating such biases is therefore an ethical necessity, even before it is a technological one. Another crucial issue concerns *the transparency and explainability* of the decision-making process: the inner workings of many AI algorithms, especially those based on deep neural networks, can be opaque to the end user, who may perceive them as a veritable "black box." This lack of visibility of the logical steps leading to the expected result hinders understanding of the decisions made by the AI agent, raising concerns in terms of accountability and the possibility of challenging outcomes perceived as unfair (Faini 2020). AI ethics requires agents to be "explainable": they must allow consumers to understand, evaluate, and control their actions. The issue of accountability and control is growing in line with the increasing spread of AI agent autonomy, illustrated, for example, by the L4 and L5 levels of user-agent autonomy proposed by Feng et al. (2025): determining who is responsible in the event of errors or damage is a highly topical legal and ethical challenge.

To resolve this dilemma, the principle of "human oversight" is invoked, which is a cornerstone of the ethical guidelines for AI in recent European regulations. Article 14 of the AI Act is dedicated to this issue and stipulates that high-risk AI systems must provide for human oversight throughout their entire life cycle. The rationale behind this provision is to prevent or at least minimize risks to the safety and fundamental rights of individuals. Control should be commensurate with the risks, which in turn are closely dependent on the level of autonomy and the context of use of the AI, and should be exercised over critical decisions made by agents, entrusting humans with control that, while not overly intrusive, allows them to:

- understand the capabilities and limitations of the AI system and monitor its operation, including how it deals with anomalies and malfunctions;
- make people aware of the phenomenon of 'information bias', i.e. the tendency to rely excessively on the output of an AI system, considering it reliable and 'neutral', especially when this tendency involves decisions that must be made by natural persons;
- exclude the AI system in particular situations or halt a procedure to ensure acceptable safety standards (Faini, 2020).

Finally, the impact on employment and society cannot be ignored: automation enabled by AI agents can have significant repercussions on the labor market: while it is true that entrusting repetitive and monotonous tasks to agents can free up resources for more creative activities, it is also true that fully exploiting this change requires a profound retraining of the workforce, implying a profound change in the concept of social equity that requires a comparison with different ways of understanding work in different global contexts (Cath et al., 2018).

The safety of AI agents is also a growing concern that emerges both from scientific literature and from the interventions of national and EU legislators. The integration of AI systems into critical infrastructures that manage large amounts of highly sensitive data makes AI agents particularly vulnerable, as they can be used for malicious attacks, compromising data integrity, privacy, and the operational functionality of systems. Among the most insidious threats are adversarial attacks, in which AI agents can be misled by subtly modified inputs that, although imperceptible to the human eye, cause the agent to make incorrect or harmf s that then propagate to the behavior of individuals (Xi et al., 2023). The robustness of agents against such manipulations is a crucial security challenge that also involves data vulnerability: the "memory" of AI agents, both short-term and long-term, and the data sets on which they are trained, can be targeted by attacks aimed at extracting sensitive information or injecting corrupt data, undermining the reliability and privacy of operations (Mittelstadt et al., 2016). System and integration security is equally critical; since AI agents interact with numerous external tools and systems (such as CRM and ERP), each connection point represents a potential vulnerability. It is therefore essential to monitor the security of APIs and communication channels in order to prevent unauthorized access or manipulation. Security can also be compromised by the risk of unexpected malfunctions: this is an aspect that must be carefully assessed because, even in the absence of malicious attacks, AI agents can exhibit unexpected behavior due to design errors, insufficient training data, or complex interactions between different system components. This scenario can lead to undesirable or harmful consequences for users, making it essential to adopt appropriate monitoring mechanisms and an effective feedback loop capable of identifying and correcting any distortions early on (Xi et al., 2023).

All these aspects require robust AI governance: this takes the form of a comprehensive framework of rules, policies, standards, and processes designed to guide the development, implementation, and responsible use of AI agents, shared as widely as possible from a supranational perspective. The primary objective of AI system governance is to maximize benefits while minimizing associated risks, with a view to aligning technological potential with the values, rights, and social objectives of different communities. Governance is supported by the existence of a robust regulatory and legal framework. At the global level, numerous countries and international organizations are developing specific laws and guidelines for AI, such as the European Union's AI Act: these regulations aim to establish clear requirements in terms of safety, transparency, non-discrimination, and accountability for AI systems, including agents (Jobin et al., 2019; Floridi et al., 2021). At the same time, the development of technical standards and certification schemes is crucial to ensure that AI agents meet predefined criteria of safety, reliability, and ethics, including the definition of metrics for measuring bias, robustness, and explainability. The reason for this dual approach is also linked to the need to manage the technological innovation process: discoveries in the field of AI are happening at a much faster pace than legislators can keep up with, who are often forced to play catch-up to ensure that legislation is consistent with the tools available on the market.

However, the need to develop robust governance models is not limited to the institutional level. Businesses must also redefine their organizational structures in order to develop clear internal policies that promote the responsible use of AI agents, providing adequate training for staff not only in the management of AI resources, but also in understanding the risks and limitations they entail. Researchers and institutions are increasingly calling for the introduction of roles and responsibilities within the organizational structure for AI management, for example by identifying and promoting those responsible for human supervision and creating channels for reporting and resolving ethical or security issues.

These aspects make AI governance not a static process, but one characterized by great dynamism: it requires regular audits and continuous assessment of the performance and impact of AI agents to ensure that they remain consistently aligned with ethical and safety objectives over time (Floridi et al., 2021). Multi-stakeholder collaboration is of strategic importance in this process: effective governance of the complex world of artificial intelligence requires synergistic interaction between governments, industry, academia, and civil society, promoting a holistic approach that values different perspectives and interests and integrates them into a common good perspective (Cath et al., 2018).

4. SOCIAL AND ORGANIZATIONAL IMPLICATIONS OF THE USE OF AI AGENTS

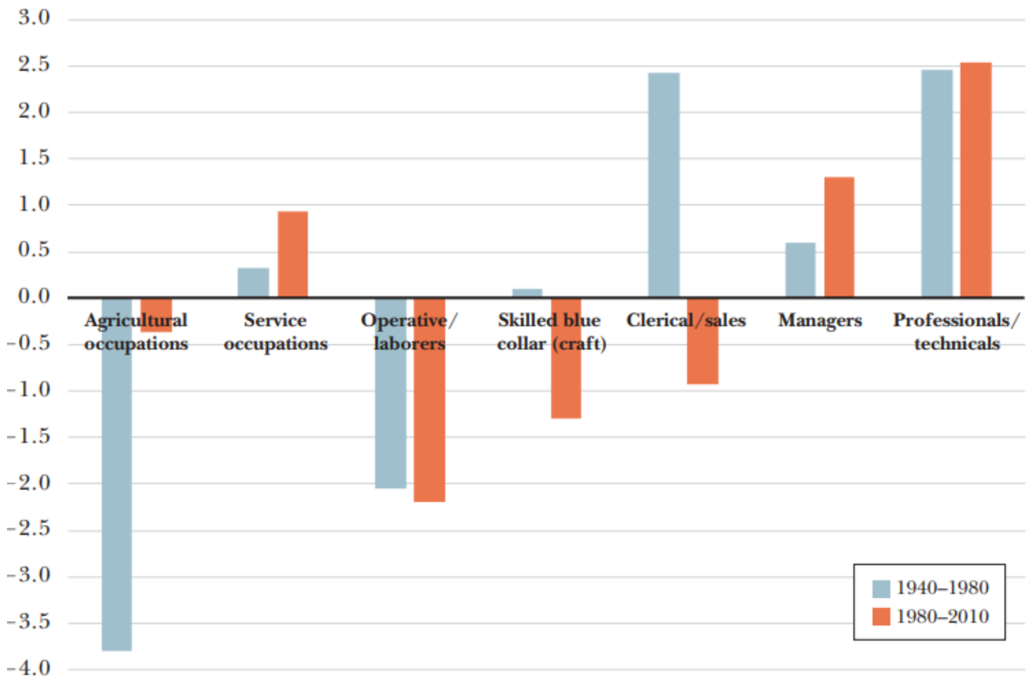
4.1 The labor market in the age of digital agents

The advent and rapid evolution of AI agents are changing many operational and relational processes and inevitably have a profound impact on the job market landscape. This is, after all, a process that has been underway for a long time: almost fifteen years have passed since two German scholars (Wolf-Dieter Lukas and Wolfgang Wahlseter) presented a report entitled "*Industrie 4.0: Mit dem Internet de Dinge auf dem Weg zur 4 insutriellen* Revolution" at the 2011 Hannover Messe. The report had a significant impact at the political, media, and scientific levels. The fourth industrial revolution heralded in the report foreshadowed an immediate future in which the Internet of Things, integrated systems, network enhancement, and, last but not least, intelligent systems would profoundly change the "game" of industrial relations and labor relations. However, it was not just a question of applying computerization to industry: this phenomenon had in fact been underway for a long time, and Lukas and Wahlster themselves recognized that the "Fourth Industrial Revolution" was in fact a revolution because it brought to fruition the insights and advances made in the previous phase, in which software and hardware were already well established in production processes (Schoning, 2018). In fact, for many years now, technologies have no longer been used solely for the management of production processes, but have become part of the value chain. Today, they enable the management (production, storage, processing, and transmission) of an enormous amount of data; the ability to evaluate (and exploit) this data within the entire value chain depends on increasingly rapid and invasive technological development. The new industrial paradigm is subjecting the world of work to a series of new challenges: new technologies require the development of new professional skills capable of understanding, managing, and, in some way, "mastering" them. At the same time, the production system has an ever-increasing need to balance the weight that the autonomy of machines (and also the transactions they create) has on human relations and changes in the world of work.

The drive towards technological innovation and the introduction of increasingly automated processes has led to profound changes in the world of work and, in , to the erosion of traditional welfare models. The introduction of AI in many productive sectors offers the possibility of reducing business costs, including (if not above all) in the area of human resource management. The automation of repetitive and routine tasks leads to a significant increase in efficiency and productivity in many sectors, but it also raises legitimate concerns about the replacement of human labor, not only for low-skilled jobs, but also in sectors that traditionally required high-level professional skills, such as complex data analysis, advanced diagnostics, or customer service. The fact that these processes involve the integration and, in some cases, complete management of processes through artificial intelligence systems, including recommendation agents, also carries the risk of reducing the capacity for inclusiveness and well-being in the world of work (Ravina-Ripoli et al., 2019).

The labor market therefore appears to be under imminent pressure to change, and the debate surrounding the introduction of AI agents into production processes has quickly heated up. According to some, the impact of AI agents on the world of work will not be such as to lead to job destruction, but rather to a polarization of skills: highly skilled activities requiring a significant interpersonal component will be protected, while routine activities will be more exposed to change. According to this interpretation, the change in the types of employment linked to the introduction of technical tools would in fact be a long-term phenomenon: studies on employment in the United States clearly show a phenomenon of declining jobs in agriculture and industry and growth in professional, technical, and managerial activities (Autor, 2015). Figure 25 highlights this phenomenon.

Fig. 25 – Changes in employment in the United States

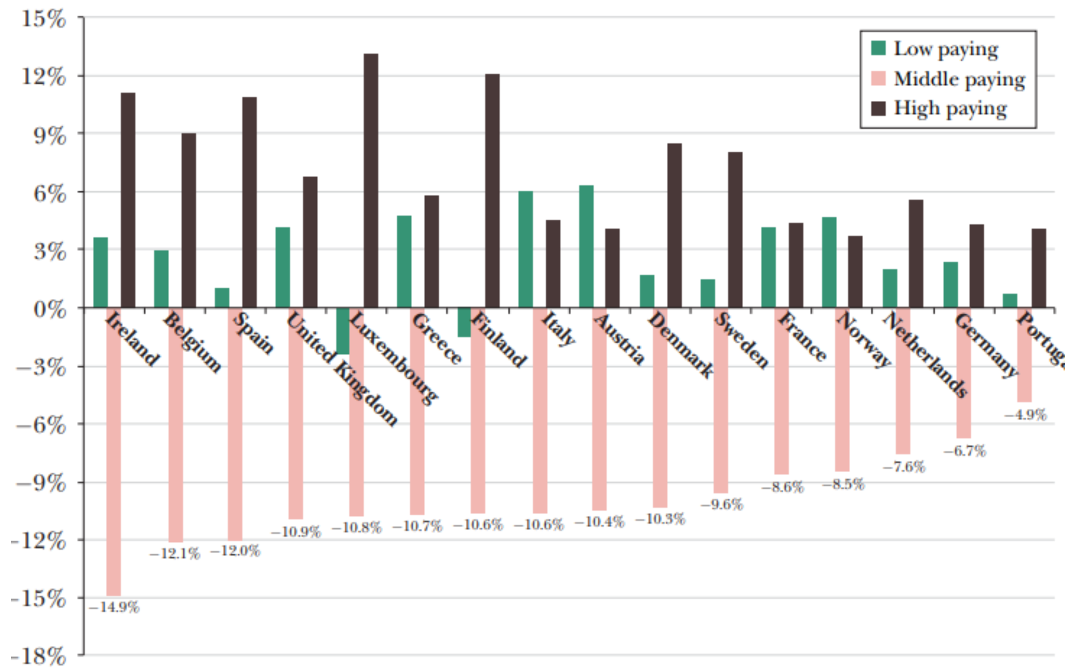


Source: Autor, 2015

The introduction of AI into the labor market would not have much impact on the number of jobs available, but rather on the quality of work. From this point of view, polarization would be linked to wage levels and would accentuate a long-standing trend that has two fundamental characteristics:

- the loss of jobs in the middle wage bracket, which is significant throughout the euro area (Fig. 26);
- an increase in jobs at the extremes (low and high wages), which, however, show a different trend depending on the country: some (Ireland, Belgium, Spain, Luxembourg, and Finland in particular) would see significant growth—between 10 and 12%—in high-wage jobs; others (Italy, Norway, Austria) are seeing growth mainly in lower-paid jobs (Autor, 2015).

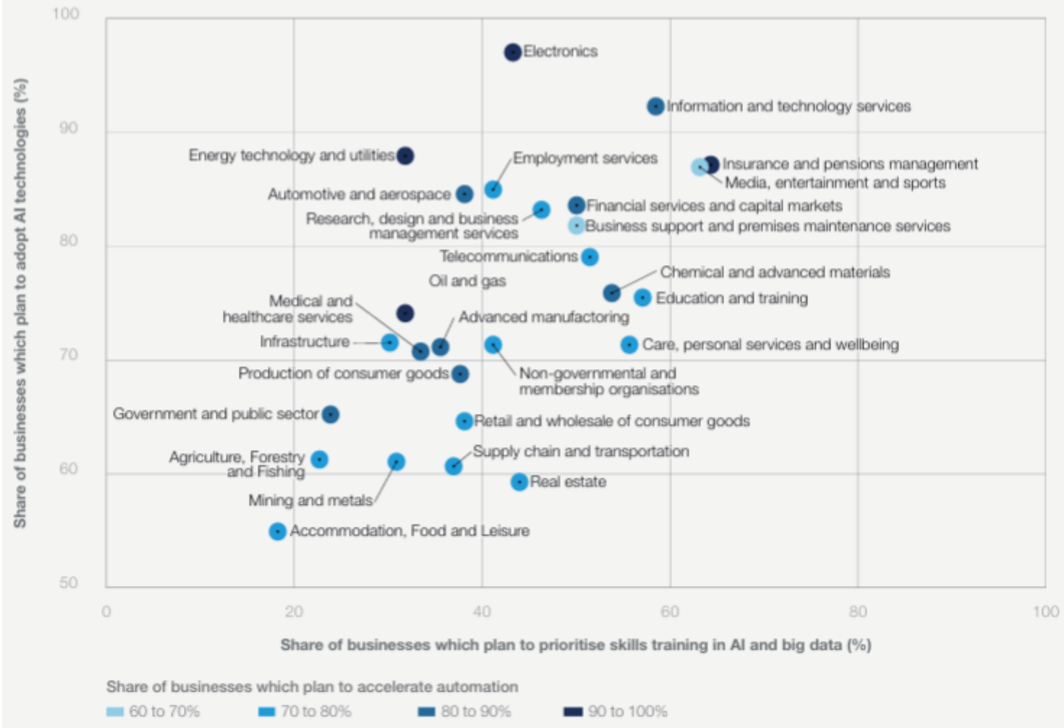
Fig. 26 – Changes in the labor market in Europe



Source: Autor, 2015

In the first chapter, we introduced a reflection on the scope of the innovations produced by the introduction of AI agents, which have all the characteristics to be considered *disruptive technologies*. Estimates of the changes that AI will bring to the labor market seem to confirm this approach. According to the World Economic Forum's *Future of Jobs Report 2023*, AI applications could lead to the loss of just under 25% of jobs globally, but create around 50% more in emerging sectors and retrained roles, indicating a net gain of 25.6%. This confirms the theoretical approach that sees interaction between humans and machines as a fundamental element of the future job market. However, the trend does not seem to be universal across all sectors: some are considered more predisposed to using artificial intelligence (energy, IT, and electronics above all), while others (accommodation, food and leisure, mining and metals, and real estate) are less likely to make a transition that involves the use of AI (Fig. 27).

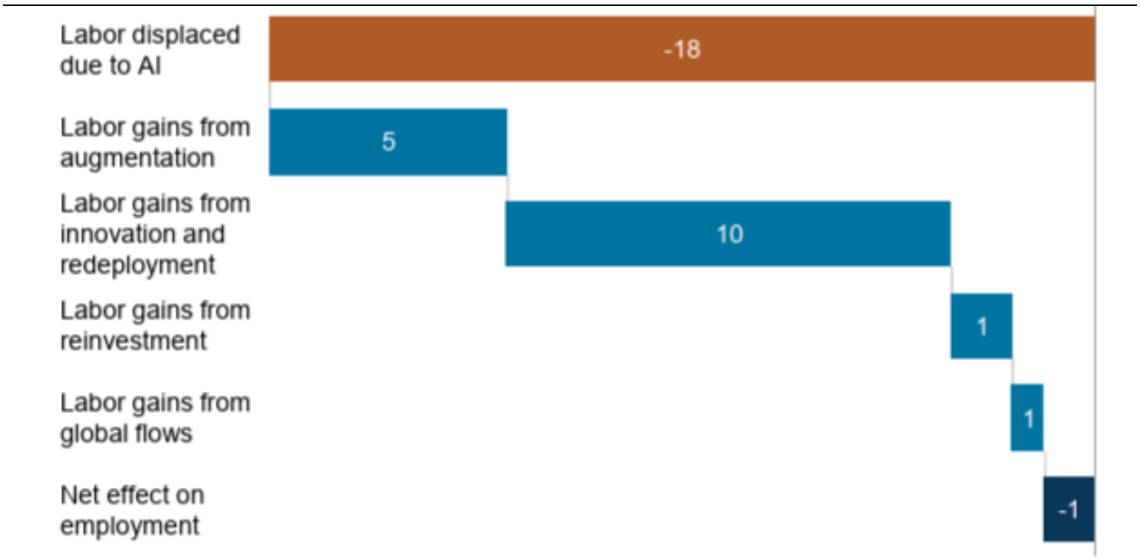
Fig. 27 – Organizations ready to use AI technology and train available human resources



Source: World Economic Forum, 2023

However, Figure 26 also highlights that not all industrial sectors ready to use AI in their production processes are equally inclined to prioritize training their staff to manage AI and big data: This is the case in the energy sector, which, despite having prospects for AI growth in the near future, does not plan to equip its staff with the appropriate skills, evidently imagining a process of labor replacement (World Economic Forum, 2023). Goldman Sachs confirms the significant changes that could affect the labor market in the near future. As many as two-thirds of jobs are exposed to some degree of automation by AI, and 25% of all jobs could be replaced (Briggs et al., 2023). The economic volume at stake is colossal: AI is expected to increase global GDP by up to 7%, corresponding to $7 trillion, by the middle of the century, with accelerated annual growth over the next two decades. This growth is driven in part by increased productivity, with industries most exposed to AI seeing three times higher revenue growth per worker (Kuzior et al, 2023). However, once again, analysts' reports show concern for the world of work: because while the trend toward polarization predicted by , with an increase in higher-paid workers, is emphasized, estimates are still uncertain in predicting the balance between jobs lost and gained. Unlike McKinsey's estimates, in fact, the impact on the world of work could be negative over the next five years (Fig. 28):

Fig. 28 – Effects of AI on the labor market by 2023



Source: McKinsey, Bughin, J. Et al., 202

As can be seen, estimates are not unanimous in indicating the impact of AI on the world of work. However, the effects of introducing AI into production processes are showing the first paradoxical examples. A striking and controversial example of this dynamic has recently emerged in the video game industry. A team of developers at a Microsoft-owned company was asked to develop an AI application that would speed up and facilitate the creation of new levels for a popular mobile game (Candy Crush). According to several specialist magazines, at least 200 people were replaced by the same artificial intelligence tools they had helped to develop (Barbera, 2025). Even a giant like Amazon seems determined to follow the same path: on June 17, 2025, Andy Jassy wrote a letter to employees in which he stated:

> "As we roll out more Generative AI and agents, it should change the way our work is done. We will need fewer people doing some of the jobs that are being done today, and more people doing other types of jobs. It's hard to know exactly where this will end up over time, but in the next few years, we expect that this will reduce our total corporate workforce as we get efficiency gains from using AI extensively

> across the company" (https://www.aboutamazon.com/news/company-news/amazon-ceo-andy-jassy-on-generative-ai).

Amazon's CEO has effectively admitted that many jobs will be lost in the coming years because AI has proven to be faster, more reliable, and above all cheaper, thus ensuring a profit for the company.

These cases raise profound questions about ethics and corporate responsibility, highlighting how, in some scenarios, AI can lead to the direct replacement of skilled roles, even those that contributed to the creation of the technology itself. Angelica Migliorisi, commenting on Andy Jassy's statement, said:

> "Jassy's statement breaks a taboo in the tech world. Until now, tech giants have tried not to say clearly that AI involves job cuts. There has always been talk of efficiency, retraining, and automation as support. But now the mask has come off. And it's not just Amazon [...] (Migliorisi, 2025).

Faced with this inevitable transformation, specialists and researchers point to reskilling and upskilling as the solution. The digital transformation linked to the introduction of AI requires people to acquire and update their skills in order to adapt to change. The two examples we have given clearly show that workers cannot compete with machines in routine and repetitive tasks and point the way to the skills of the future: data management, critical evaluation of information sources, and the ability to integrate new digital solutions into organizational processes will enhance the complementarity between humans and AI. This implies a marked shift in the emphasis of training: skills such as critical thinking, creativity, complex problem solving, emotional intelligence, negotiation, and the ability to work in interdisciplinary teams are skills that AI, however advanced, still struggles to replicate fully. However, to enhance these professional skills, it is essential to develop digital and technical skills that enable workers to interact, supervise, and collaborate effectively with AI agents (Kumar et al., 2024).

4.2 Leadership and business organization: how governance is changing in the digital economy

The widespread introduction of AI agents is not limited to changing the way individuals approach work, nor can it be confined to the redefinition of work processes: among the most significant consequences of automation processes through artificial intelligence algorithms is the profound transformation of organizational structures and corporate leadership models. Traditional hierarchies, often based on rigid chains of command and centralized decision-making processes, are being challenged by the emergence of a new paradigm in which information and analytical capabilities are distributed across all levels of the organizational structure and amplified by artificial intelligence: AI agents are redefining not only operational responsibilities but also the decision-making chain, making it less vertical and more dynamic. Even decisions, once the exclusive preserve of managers and executives, can now be supported by sophisticated algorithms capable of processing historical data, market trends, and external factors, running simulations more effectively and quickly than human intellect could ever hope to achieve. Taken to the extreme, this phenomenon could even entrust the decision-making process entirely to a suitably trained algorithm: AI agents based on LLMs are equipped with the ability to continuously learn and adapt to the environment in which they operate and interact, and thanks to their ability to process data, they are able to ensure a faster and more effective response to change. They also have another advantage over traditional decision-making methods: they can apply predictive models capable of identifying current trends and preparing the organization for imminent change (Huang, 2023).

The integration of AI is fostering the emergence of new organizational models that the literature refers to as AI-powered organizations. The way in which AI agents are changing organizations is something more profound than greater efficiency in production processes or increased productivity and competitiveness in the market. Rather, they enable AI to be integrated into every level of the business organization, starting with decision-making processes and redefining business models and corporate culture. An organization that knows how to integrate all the possibilities offered by AI into its decision-making processes and production s radically changes the way it reacts to the market: it can abandon a purely reactive approach, based on the ability to react to market changes, in favor of strategies that anticipate the future, thanks to the predictive capabilities of AI agents. Corporate governance is thus significantly impacted: organizations are called upon to develop new capabilities for supervising and managing AI agents, ensuring that they operate in line with the company's strategic objectives, ethical values, corporate vision, and, ultimately, applicable regulations. This includes defining clear frameworks of responsibility for decisions made by algorithms, the creation of auditability mechanisms (i.e., traceability and verifiability of decisions made with the help of AI or even autonomously by AI agents), and the development of internal skills for monitoring performance and mitigating risks associated with AI, such as the aforementioned risk of algorithmic bias or unethical drifts (Mantymaki, 2022).

The literature has identified some critical issues in managing the transition to an AI-powered organization. The first is the lack of coordination with which AI has often been introduced into companies: this has led to fragmentation that limits the interaction that AI can have with governance processes, both in relation to the "domain" of information technology and in terms of interaction with corporate governance as a whole. A second problem concerns the ethical sphere: partly as a result of the fragmentation mentioned above, the search for the right balance between optimizing automation and maintaining adequate human supervision to limit the risk of delegating too much judgment and responsibility to autonomous systems is often limited. This area, however, represents an important challenge involving multiple levels of the organizational structure. From a technical point of view, the challenge is primarily linked to data quality and the possibility of integrating AI with existing systems within the company: this requires highly specialized talent and implies a redefinition of the flow of information within the organization. Two other significant dimensions fall within the scope of considerations involving the ethical aspects of business conduct: once algorithms have been trained appropriately and the possibility of human decision-making bias has been limited, it is necessary to address the challenges related to data security and the protection of confidential information, and to explore issues related to changes in the labor market.

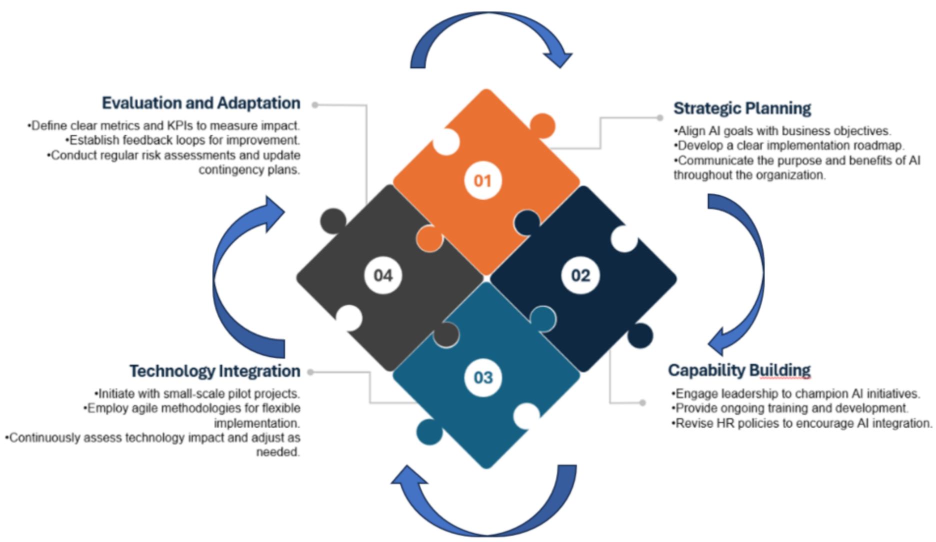
AI is introducing important factors of change in the organizational structures of companies. However, change must be guided. The literature distinguishes between two types of change: the first is what is defined as "accidental change" (Burke, 2002), which is characterized as a form of reaction to external events that are unpredictable and beyond one's control. For this type of change, planning is virtually impossible, and the only possible response is to create a resilient organization capable of absorbing the shock, minimizing damage, and, if possible, maximizing the positive effects. The change planned by organizations is very different. In this case, change arises not so much from the need to deal with a crisis, but from the realization of the gap between the desired state of performance (a certain level of market penetration; customer satisfaction and loyalty; profits generated, etc.) and the actual state. In change initiatives, top management will need to make decisions using well-defined and structured techniques and tools, and this is precisely where AI agents come in. Facing change, in the dual perspective we have introduced, is no longer an extraordinary situation, but an ordinary aspect of management. However, every change must be managed, and this requires a pragmatic approach to what organizational science defines as "Organizational Change Management." This is a fascinating and complex discipline that deals with the set of techniques used to manage change and address the challenges that arise in the ecosystem in which organizations operate. In the case of companies wishing to integrate AI agents, change is a four-part process:

- Strategic planning. The decision to align AI objectives should be developed through the definition of a road map that sets out budgets and important organizational and management steps. These also include the need to communicate the goals and benefits that AI can bring to the company: a step that is not of secondary importance, given the concerns that the introduction of AI in the workplace raises among professionals;
- Capability building. Every change is above all a cultural phenomenon. In the case of AI, addressing change means working on a very complex level, because it involves the participation of leaders (formal or informal) so that they can act as drivers to encourage the integration of AI into all business processes, including human resource management policies.
- Technology Integration. This means continuously assessing the impact of AI on information technology and deciding on the level of integration. What emerges from the literature, in fact, is a partial integration of AI into business systems. However, this means losing two of the main advantages of information systems in general and AI agents in particular: the ability to adapt to the contexts in which they find themselves and to be scalable. For this reason, the literature suggests that the most effective strategy for defining an AI-Organization is to plan for its gradual introduction into operational, management, and decision-making processes, starting with small pilot projects.

- Evaluation and Adaptation. The change brought about by AI Agents is subject to two important dynamics for the organization: it moves progressively in breadth across a growing number of functions (defined by the organization during the planning phase, but also driven by increasingly rapid technological change) and in depth, restructuring management rules, corporate values, and corporate culture. For this reason, it is important not only to define metrics and KPIs that measure the impact of artificial intelligence on financial performance, but also to establish assessments that can oversee its use through risk assessment procedures and the updating of contingency plans (Nayana et al., 2024).

AI agents thus become not mere tools for technological innovation, but a fundamental strategic component of organizations, affecting both the professional and human dimensions of individuals and, ultimately, the life of the entire organization. For this reason, it is important that the phases we have described are included in a circular process that keeps the attention of corporate leadership and th s focused on the operational and ethical issues related to the introduction of AI agents into business dynamics (Fig. 29).

Fig. 29 – Change management in AI-Powered organizations



Source: Nayana et al., 2024

The consequences of these new organizational structures also extend to leadership. The new leadership emerging in this context can no longer rely solely on the authority derived from hierarchical position, but must be based above all on the ability to interpret and act on the insights generated by AI agents. The role of the leader is therefore called upon to evolve from that of a mere controller to take on the characteristics of a facilitator, capable of interpreting the complex dynamics that emerge from the interaction between human and artificial intelligence. Organizations that choose to integrate AI agents as a strategic component of their business implicitly accept a challenge that involves not only individual workers and teams as a whole, but also the way leadership is understood.

Defining leadership has never been an easy task: the way in which the role and functions of leaders are conceived has been subject to constant review in light of production systems, the consideration of work determined by social relations inside and outside the workplace, the regulatory system, and other factors involving a multitude of cultural dimensions. The integration of AI agents ( ) into organizations involves some important aspects related to leadership that emerge from an analysis of the literature and that are worth highlighting. The first starts from the consideration that being a leader is not the same as being a manager. Although the terms "manager" and "leader" are often used interchangeably, they are not the same and do not necessarily coexist in the same person. The work of a manager is much more reminiscent of the characteristics of a "boss" as understood in the literature of the 1920s, starting with Max Weber's reflection on power. The following table identifies the characteristics that distinguish these two figures:

Table 2 – Characteristics of leaders and managers

Leader	**Manager**
Innovation	Administration
Development	Maintenance
Inspiration	Control
Log-term vision	Short-term vision
Allocation of responsability to staff	Allocation of task to staff
Challenging the status quo	Tacit acceptance of status quo

Source: Flores, 2008

A manager, therefore, does not necessarily possess the characteristics and skills to be considered a leader. However, these characteristics refer to traditional professional contexts, to the point that it is worth asking whether they are still relevant and effective in contexts such as the current one, characterized by the need to integrate AI agents into organizational structures. The changes in the world of work that we have described can create and amplify a series of tensions towards a complexity in which change and the centrality of the individual seem destined to take on new meanings, but no less important than in the past for ensuring the survival of organizations. This challenge does not seem to be able to be addressed with the same "qualities" of leadership that were essentially valid in the past, not even with the practicality of a manager. Ultimately, it may depend on a cultural approach that contrasts the propensity to innovate with the need to ensure the continuity of the organization.

The environmental and organizational complexity of the contemporary world requires that management behaviors now focus more on the deep involvement of people: AI agents have created uncertainty among workers, who have seen the rules, habits, and very "rituals" with which they approached the world of work in the past turned upside down. In order to cope with change, and change on the scale described in the previous pages, people need to be able to count on leaders who are capable of being innovative and able to motivate their teams, activating a series of reinforcements that guide people to get back into the game by proactively and continuously retraining and updating their professional skills.

The "new" leaders in AI-powered organizations will therefore need to master several domains:

- Cognitive amplification domain. This emerges from distributed cognition theory, according to which the set of thought processes, learning, memory, and problem-solving skills are not the property of individuals but become the heritage of the organization as a whole, also involving tools (including technological ones), the physical environment, and the social ecosystem. AI systems amplify this dynamic by distributing the cognitive load between human and technological agents. In this perspective, leaders are called upon to develop skills that enable them to deepen data analysis, recognize patterns and deeper knowledge, and 'train' people in critical thinking.
- Collaborative learning domain. In organizations, knowledge will increasingly be the result of collaboration between AI agents and humans. This means that leaders

will need to promote the ability of humans to interact with AI agents. This is not just a matter of technological application: rather, it appears necessary to guide behaviors that lead to a new understanding of roles and skills within the organization.

- Ethical governance domain. Leaders' actions must always be aligned with values, people's rights, and organizational objectives. In this perspective, they must be able to address the ethical challenges that emerge from the current context, in which frequent and intense interaction between individuals, AI agents, stakeholders, and consumers takes on a new significance compared to the past.
- Adaptive evolution domain. Leaders must be able to manage change, which, especially when it involves technology, is increasingly fast, continuous, and sometimes unpredictable. In complex organizations characterized by interactions between AI agents and humans, emergency management, self-organization, and feedback monitoring are essential, considering that AI systems are capable of evolving over time in response to environmental stimuli and increasing technological capabilities (Sposato et al., 2025).

The adoption of AI agents is catalyzing the emergence of new business models and the redefinition of existing ones. Predictive supply chain optimization that drastically reduces process costs and times, predictive maintenance that prevents costly failures, and automated customer services that improve the user experience and free up human resources for more complex tasks are just a few examples of the strong integration between AI agents and humans. Sectors such as finance, healthcare, logistics, and retail are already undergoing profound transformations, with AI enabling mass personalization, predictive consumer behavior analysis, and optimization of operations on an unprecedented scale. The ability to leverage data and AI to create value is now a critical success factor and a necessary driver of competitiveness for businesses. However, the strategic management of AI agents also depends on the degree to which leaders are able to adapt more quickly to ever-changing market dynamics.

4.3 The social implications of the widespread adoption of AI Agents

According to some literature, entrusting decision-making to AI agents has an additional advantage. It could limit the effects of human cognitive biases; however, it introduces the issue of algorithmic biases, which occur when training data is flawed or biased, or when decision-making logic is not transparent. The issue of algorithmic bias is particularly interesting. Its roots can be traced back to social phenomena such as discrimination, injustice, and social inequality: for a system to be considered fair and impartial, it must be able to ensure the balanced distribution of benefits and burdens among those who are part of it. Algorithmic bias occurs when an AI algorithm favors (or harms) certain individuals or groups without there being a reasonable, objective, or legitimate reason to support this inequality. The consequences of algorithmic bias manifest themselves on several levels. Individual consequences include higher prices for customers and employment inequalities that affect minorities in the workplace. Organizational effects can include violations of equal opportunity policies, the creation of an unethical climate within the organization, resulting in higher employee turnover and high customer churn due to algorithmic discrimination and dissatisfaction (Kordzadeh et al., 2021). Social effects include an increase in the wealth gap between historically disadvantaged groups and others. There is no shortage of examples of this in the news. For example, the algorithm used by Apple Card to decide credit limits has been accused of assigning much lower credit limits to women than to their spouses, even if the wife has a higher credit score (Thorbecke, 2019). Another very different case relates to the application of algorithms in the judicial sector: assessments carried out using AI algorithms can evaluate the risk that an individual may reoffend, thereby avoiding mass incarceration while ensuring optimal levels of public safety. However, it has been highlighted that algorithms could emphasize inequalities in the way minorities are treated: A 2018 study in the United States demonstrated the existence of a bias linked to belonging to different communities: the risk of recidivism was found to be overestimated for the black population and in the Hispanic community (Hamilton, 2018).

These are risks that regulators and legislators are aware of. Measures have been taken to combat bias based on data analysis. For example, the European Union's General Data Protection Regulation (GDPR) and the US Future of AI Act of 2017 impose restrictions on data processing and commercial practices based on artificial intelligence, in the hope of promoting algorithmic accountability and reducing algorithmic bias. However, this is tricky: in highly innovative fields, regulation lags behind technology and can't always offer policies and legal tools to mitigate bias in algorithmic practices. Plus, it's still unclear how interactions between people and algorithmic systems can shape, trigger, or prevent data-based bias in organizational decision-making.

The large-scale adoption of AI agents transcends economic and organizational spheres to profoundly impact the social fabric, raising complex questions, particularly about privacy, surveillance, and the protection of fundamental rights. The inherent ability of these systems to collect, process, and correlate immense volumes of personal data—often without the full awareness or consent of individuals—poses serious threats to the protection of individual privacy. From facial recognition systems used in public spaces to predictive analysis of online behavior for commercial or control purposes, the potential risk of pervasive surveillance in individuals' lives is no longer just a theoretical possibility, but a real one. The profiling and widespread monitoring capabilities exercised through AI agents raise serious concerns about the possibility of influencing individual choices, restricting civil liberties, and creating a sense of constant observation, undermining trust in institutions and in the very ability of AI tools to improve people's lives (Kordzadeh et al., 2021).

We have seen how institutions are trying to equip themselves to limit the risks associated with the increasingly pervasive use of AI agents. Regulation EU/2024/1689, known as the AI Act, is one of the first global attempts to regulate artificial intelligence in a systematic manner, seeking to make systems developed and used in the EU context safe, ethical, and respectful of citizens' fundamental rights and EU principles.

The European legislator's approach is risk-based. The regulation defines three levels of risk, the highest of which (defined as unacceptable) leads to a ban on the use of a number of systems in Europe. In particular, the following are prohibited:

- systems that pursue social scoring by governments or companies;
- the use of manipulative techniques capable of influencing people's behavior;
- systems that exploit the vulnerabilities of specific groups, amplifying social differences against minors, disabled people, minorities, etc.;
- systems that identify people in real time in public spaces based on biometric data. This possibility is permitted in limited cases and only for reasons of public safety;
- emotion recognition.

The AI Act assigns a high risk to AI systems used in critical infrastructure, human resource management, healthcare treatment and diagnosis, and in the judicial and educational fields. According to the European Union's approach, these systems must be carefully monitored because, in the event of malfunction or bias, they could cause significant damage to people's fundamental rights, safety, and health. Developers and users are therefore required to comply with stringent requirements and constraints, one of which is to ensure effective human oversight.

AI agents generally fall within the category of low-risk applications, but users need to be aware that they are interacting with an artificial intelligence system.

The European Union's approach is an attempt to balance the need to protect society from the potential risks of AI systems with the goal of promoting innovation. However, the *ex ante* risk-based approach of the law could lead to regulatory obsolescence: every technological innovation forces the legislator to reopen the regulatory process and entails the risk of insufficiently classifying AI applications into different risk categories, overestimating or underestimating the real risk. Although European law and regulatory procedures provide important safeguards through the involvement of multiple actors and technical bodies, their actual ability to identify and correct the unintended consequences of AI use appears rather limited. This is due, in general terms, to confirmation bias and automation, which lead people to place excessive reliance on the results of AI systems without critically questioning their validity (Rangone et al., 2025).

The application of these regulations involves a number of complexities. One of the most significant lies in identifying legal responsibility in the event of errors or damage caused by autonomous AI agents—is it the designer, the developer, the operator, or the company that implements it? From this point of view, the regulatory and conceptual framework still appears rather inadequate to define reality.

Finally, the proliferation of AI agents raises issues that go beyond mere regulatory compliance, touching on social cohesion and fundamental rights. The ability of algorithms to create "filter bubbles" and "echo chambers," amplifying misinformation and polarizing opinions, poses a threat to democratic stability and civil dialogue. Unequal access to AI-enabled technologies, or their implementation without considering the needs of minorities or vulnerable groups, can exacerbate social and economic inequalities (Rangone, 2025).

5. COMPARATIVE CASE STUDIES

We have seen how the evolution of AI agents, from simple chatbots to complex autonomous and predictive systems, is reshaping corporate organizational architectures and business models. This chapter aims to examine, through a number of case studies, the practical application of these technologies in different operating sectors. The objective of this analysis is twofold: on the one hand, it seeks to identify the technical and strategic specificities adopted by leading companies in their respective sectors with regard to the strategy for using AI agents in different core business processes; on the other hand, the analysis aims to introduce some elements of comparison which, despite the diversity of the sectors of use and the strategies adopted by companies, allow for an evaluation of the use of AI agents in terms of operational efficiency, organizational impact, and effects on business performance, providing an empirical framework to support the arguments developed in the previous chapters. Finally, the comparison between different operational strategies will highlight the potential and limitations of AI agents.

5.1 AI Agents for customer service: the case of Klarna

The first case study concerns the application of AI agents in customer service. Managing customer and consumer relations through intelligent chatbots and virtual assistants is one of the most significant areas of application for AI agents in modern organizations and has radically changed the way companies interact with their customers. On the one hand, AI agents allow users to receive timely responses without time restrictions; on the other hand, they allow for a high degree of personalization in responses, thanks to the ability of AI agents to adapt to different communication channels and eliminate communication barriers, especially linguistic ones.

On the other hand, however, limiting the scope of AI agents to direct user relations would be restrictive. They also have a positive impact within the organization: by automating highly repetitive tasks, they free up resources and skills for activities where a human approach can bring greater added value than an algorithm. Furthermore, the collection of real-time insights builds a solid information system that is disseminated throughout the organization, enabling the company to acquire information and generate internal knowledge which, when properly directed, can lead to continuous improvement not only in the relationship between the company and the user, but also within the company itself (Mathew, 2025).

The case we have chosen to analyze to highlight the importance of an effective strategy in managing AI agents in customer relations is that of Klarna, the Swedish fintech company founded in 2005 whose core business is financial services in the "Buy Now, Pay Later" (BNPL) sector. Essentially, Klarna offers its users the ability to shop online and in-store, providing flexible payment solutions through an extremely user-friendly interface and an effective and highly regarded customer support system. The centrality of the customer experience—driven by significant technological investments—is one of the company's strengths.

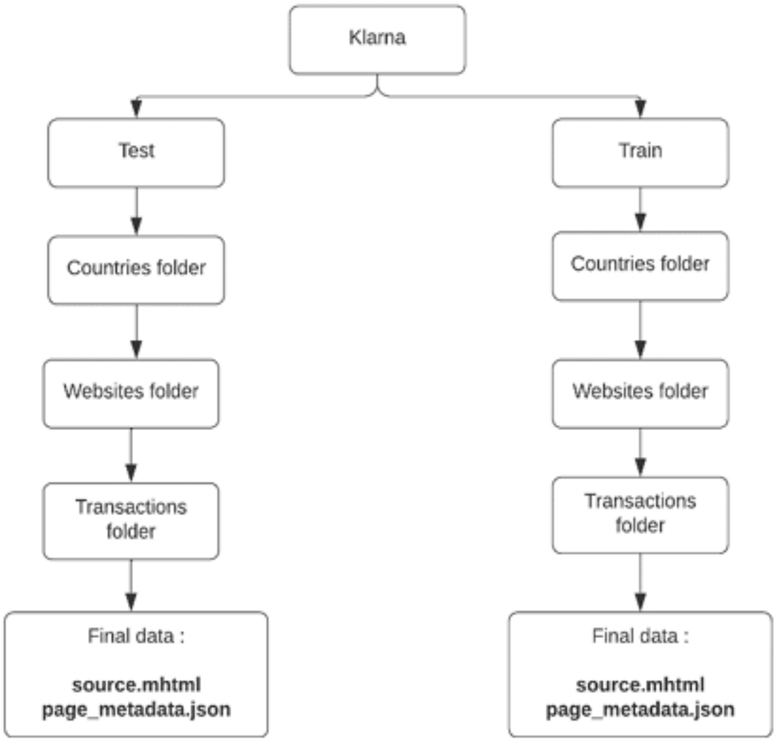
In 2023, Klarna announced that it had replaced approximately 700 employees (approximately 40% of its workforce) with AI agents. The decision was motivated by three key reasons:

1) the need to reduce costs through a significant reduction in personnel-related operating expenses;
2) increased efficiency, as AI agents were expected to improve response times and lead to faster problem resolution than human operators;
3) the scalability of the systems, which, thanks to their flexibility, can adapt to customer call flows.

The AI Agent developed by Klarna has a modular structure and integrates supervised machine learning models, whose structure is based on an Imitation Learning model trained on data in order to learn effective decision-making strategies from interaction with consumers. The training phase allows the AI Agent to replicate optimal behaviors during customer interactions in the operational phase, increasing the accuracy and consistency of responses over time thanks to the information learned from interactions with customers themselves. From an operational point of view, the system is set up to be " ," substantially limiting random responses and ensuring greater reliability and repeatability of results.

Klarna provides a public dataset in which each product is labeled with five key pieces of information (price, image, name, add to cart button, and cart). The data can be aggregated from a variety of different sectors and outputs a large amount of digital data from many channels. The dataset has a tree structure with the root "klarna" at the highest level. The algorithm's operating structure involves parallelism between test data (files with the extension .json) and training data (files with the extension .mhtml). (fig. 30; El Attaoui et al., 2023).

Fig. 30 – Structure of the Klarna dataset



Source: El Attaoui, 2023

From a process perspective, the AI Agent algorithm performs a series of structured tasks:

1. It reads all the folders contained in the tree structure and searches for specific files containing the labeled data, saving them in memory;
2. All information contained in the data files is retrieved, while searching for specific patterns or markers related to relevant information. This information is saved in a file called "database.json." The technical characteristics of the file make it particularly easy for the machine to process. The database.json file stores all the information collected, making it accessible for further processing.
3. Data extraction is performed through an API that converts the data, making it more structured and usable for further processing.

Thanks to continuous learning mechanisms, the system is able to modify its behavior based on new requirements that are progressively recorded based on both interactions with users and changes in business processes, thus increasing the accuracy, relevance, and impartiality of responses (El Attaoui, 2023).

The application of the AI Agent for customer relationship management has had several implications for Klarna. These can be summarized as follows:

1. implications for the world of work. The decision to reduce the workforce by 700 employees represents a real replacement of staff. Although Klarna has stated that it has offered many of them the opportunity to retrain within the company, this strategy has sparked some debate in public opinion because it has highlighted the implications that AI can have on the world of work in terms of net job losses. There has been no shortage of controversy and negative comments about what is considered an overly aggressive approach by the company.
2. Customer experience. According to the data, Klarna's AI agents have reduced waiting times by 25% (from 11 to 2 minutes) and improved problem resolution by 30%. Overall, 2.3 million conversations are handled. The AI agent is available 24 hours a day and "speaks" 35 languages. Its performance has been judged superior to that of the response model carried out solely by human operators.
3. Financial implications. Klarna has predicted a $40 million increase in profits during the first year of implementation of the new system with AI Agents. The positive effects in terms of financial performance were already evident in the short term (AIleaders, 2024).

It would therefore seem that Klarna's is a success story (albeit marred by the dismissal of hundreds of people) in the application of AI agents in customer service. Klarna seemed capable of developing a successful strategy, which earned the company impressive scores in international evaluations and a series of accolades from trade magazines and observers, who saw the 'Klarna case' as a benchmark for the application of AI agents in customer relations. However, the story told so far only applies to the period immediately following the implementation of Klarna's strategy: in the medium term, however, things began to change. Users began to complain about overly stereotypical responses: when put to the test, the AI agent proved capable of handling questions excellently, but was much less effective in managing the relationship. In the medium term, customer loyalty declined and many left the platform. As a result, Klarna resumed hiring staff and reassured its customers that they would always have a human being available to solve their problems. The Klarna case highlights that cost reduction achieved through simply replacing human operators with AI is, in the context of customer experience, almost entirely illusory. The savings achieved by 'cutting' staff risk being offset by the loss of customers: although less obvious (and less immediate) in terms of financial performance, there is no doubt that the economic damage is considerable: the loss of a customer who is annoyed and disappointed by a frustrating experience is a huge cost for the company, because it forces it to invest in acquiring new customers: an effort that can lead to a cost increase of up to seven times greater (Gallo, 2014). It should also be noted that Klarna is not the only example of a rethink of a strategy of robust AI injection into user relations. As many as 55% of companies that 'cut' employees to introduce AI agents have had second thoughts and have returned to hiring human staff in an attempt to regain the level of empathy that AI agents do not yet seem able to fully provide (Martinez, 2025).

The Klarna case seems to demonstrate the need to take a less absolute approach: AI should be used with precision, leaving repetitive tasks to it and entrusting humans with all those cases where empathy, critical thinking, and the ability to analyze and make decisions on a case-by-case basis still constitute added value in the relationship with the consumer (Haun, 2025).

5.2 AI agents in project management: AutoGPT

The ability of AI agents to analyze large volumes of data from different sources is one of the characteristics that has led to their adoption in project management. The Association for Project Management defines Project Management as

> "[...] the application of processes, methods, skills, knowledge, and experience to achieve specific project objectives according to the project acceptance criteria within agreed parameters. Project management has final deliverables that are constrained to a finite timescale and budget."

Each project therefore has an objective, which must be achieved by adopting a series of activities and tasks, in accordance with a schedule defined by deadlines and the integrated management of resources identified within a budget (Fig. 31).

Fig. 31 – Project Management Model

Source: Personal Draft

Project management is an operation that can be highly complex: it requires the acquisition and use of a large amount of information; it makes it necessary to monitor the progress of activities in order to verify their impact (actual and future) on the expected results; it requires constant risk assessment in order to implement contingency plans in the event of events or situations that could compromise the final result. Finally, it is worth noting that projects, especially large-scale ones, often involve multiple organizations, each with its own organization, processes, and data that, at least in part and for a limited time, must be shared with other structures.

The integration of AI agents into project management can be a strategic element in supporting the decision-making process. The adoption of AI solutions in project management is based on the ability of agents to interact with existing systems, automate repetitive tasks, and support more complex activities through the integrated management of large amounts of data and the provision of contextualized recommendations.

AutoGpt is an autonomous agent belonging to the Generative AI family and is based on an LLM architecture with memory, planning, and dynamic interaction mechanisms with external agents. This approach allows AutoGpt to be used in project management to tackle complex tasks that require sequential reasoning, contextual knowledge management, and constant interaction with APIs and business systems, in order to automate constant and repetitive operations as much as possible. Key elements of AutoGpt are:

- memory management, thanks to backend tools that allow the AI Agent to store information processed during specific tasks and reuse it later, effectively retrieving data and ensuring a self-learning dynamic;
- the management of explicit Chain of Thought (CoT). The management of these chains is fundamental in project management because it allows complex problems to be broken down into logical intermediate steps, thus assigning activities and tasks according to the project plan. In this way, not only are logical steps and the transparency of the decision-making process improved, but it also facilitates the monitoring of the project activities themselves;
- the adoption of advanced prompting techniques and autonomous decoding mechanisms that allow effective and reliable responses to be developed even in scenarios characterized by multi-step planning or where the ability to adapt to constantly changing contexts is required (Yang et al., 2023)

In the context of project management, AutoGPT can be used for numerous applications:

- automatic or semi-automatic report generation;
- monitoring of work progress;
- dynamic resource allocation (budget management);
- predictive scenario planning to verify the impact of actions taken and, if necessary, adopt contingency plans in case of deviations.

The competitive advantage offered by AutoGpt is its ability to interact with existing business systems, retrieving information from CRMs or other platforms already in use by individual project participants. This aspect is strategic in project management, especially when projects involve different partners, each of which produces information in an autonomous galaxy of systems that should be integrated into a single management tool. AutoGpt has the ability to integrate these systems, as well as to acquire information from digital sensors or Internet of Things systems, ensuring real-time monitoring of project performance (Korzynski et al., 2022).

Like any Generative AI system, the process of collecting, organizing, and managing data is in fact a tool for analyzing and disseminating knowledge within an organization engaged in project management. The great versatility of the algorithm that governs the AI Agent allows it to be applied at different project levels:

The literature has highlighted how, in project management, the use of AI Agents such as AutoGpt is related to the way in which organizations adopt and deal with paradox theory. Modern economies are subject to a series of pressures involving a multitude of dimensions which, compared to the past, highlight a growing complexity and speed of change. This confrontation implies a daily effort to adapt to new market dynamics, to modify production processes, and to introduce innovations that often risk upsetting established internal balances. These often unpredictable dynamics significantly affect decision-making processes, which are now more than ever subject to a series of conflicting forces. Even the way organizations make decisions must adapt to the complexity and speed of change. Every company faces dilemmas and problems on a daily basis, making choices that can potentially fuel tensions that reverberate throughout the entire organizational structure and are at the root of contradictions and paradoxes that involve not only managers but also all professional levels throughout the organization (Putnam, 2015). The moment of choice is as essential as it is critical for every organization: it involves a comparison of different ideas and perspectives that give rise to a strong internal dialectic, highlighting contrasts and contradictions and often leading to polarization around conflicting, if not irreconcilable, positions. The reason why this dynamic creates anxiety and stress in organizations (and naturally also in the people who are part of them) is linked to the perception that the moment of choice is determined by a clash of incompatible visions: a clash that feeds what doctrine refers to as contradictions within organizations.

The term *contradiction* refers to a polarization (of processes, values, and perspectives) that develops in two ways: on the one hand, each position defines itself as radically different from the other, identifying distinctive characteristics almost "by contrast"; once defined, the positions then tend to negate each other (Putnam, 1986). This dynamic creates a binary culture in which any attempt to make one position prevail ends up—even when successful—defining (by contrast) and polarizing the alternative position, which,

even when it succumbs at the moment of choice, manages to represent cultural and social demands, power systems, and beliefs that persist within the organization.
Contradictions are now considered an intrinsic component of organizations, and it follows that the ability to manage them is an essential element for any organization. Each stage of the decision-making process involves a redefinition of the system and of the internal relationships with n within the organization, as well as a willingness on the part of all actors to question themselves. Tensions can also directly affect the professional lives of each of them, making it plausible that the resulting contradictions lead to the mutually exclusive polarizations we have mentioned. However, it is not necessarily the case that different positions are always irreconcilable: rather, they are one of the manifestations of complexity, and in the current economic landscape, companies are called upon to manage contradictions in order to gain competitive advantages and ensure long-term sustainability. This, however, generates organizational paradoxes, or

> "[...] contradictory yet interrelated elements that exist simultaneously and persist over time" (Smith et al., 2011, p. 382).

Organizational paradoxes are not exceptional situations for organizations: they are present in all organizations and cannot be avoided. This requires new skills, in particular what some literature refers to as *organizational ambidexterity*. This implies the ability to manage

> "[...] simultaneous, yet contradictory, knowledge management processes, exploiting current competencies and exploring new domains with equal dexterity" (Andriopoulos et al., 2009, p. 696).

If left unmanaged, organizational paradoxes can cause enormous problems for organizations because they can undermine individual self-efficacy and well-being and reduce operational capabilities. Managing paradoxes requires corporate management to internalize the assumption that contradictions must be addressed starting from the ability of opposites to recognize and legitimize each other: by 'institutionalising' the presence of tensions and recognising the value of each position, the disruptive capacity of polarisation is reduced, acknowledging that every idea, every construct and every vision can coexist and contribute to the well-being and success of the organisation as a whole.

According to paradox theory, tensions within an organization cannot be resolved; it does not deny the presence of tensions that lead to internal competition; but unlike classical theory, these tensions are inter , interconnected, and interdependent, and are an intrinsic characteristic of organizations, present at all levels of the structure and persistent over time.

Paradox theory dates back to a complex theory based on the analysis of dialectics within organizations. AI agents can manage the forces that arise along three main axes more effectively than humans can:

- **Opposition**: managing paradoxes involves the ability to coordinate opposing organizational elements. These are dimensions present in the organization that, when isolated, have their own logical consistency but become irrational when attempts are made to combine them. While classical theory proposed choosing one option and rejecting the other, paradox theory requires reaching a synthesis that reconciles both.
- **Interdependence**: however opposed they may be, these objects are, as we have seen, interconnected. The existence of one is due precisely to the existence of the other; the absence of one of the two opposing terms does not determine the end of the paradox, but rather the search for a further 'opposite' element on which to build identity and contrast;
- **Persistence**: the tensions that arise between opposing and interdependent elements cannot be resolved definitively: they persist over time and, if not managed, can lead to disruption and conflict (Smith et al., 2011).

The use of AutoGpt (like other AI agents) is not, in this context, only related to the management of repetitive and standardized tasks. Some recent positions in the literature have highlighted how AI can also be applied to the process defined as "augmentation," i.e., the use of AI to increase the potential and capabilities of human beings rather than replace them. This shift from the automation of repetitive tasks to proactive support is one of the most interesting perspectives found in the literature and involves a broader application of project management processes through the use of generative AI (Raisch et al., 2021). This is a very different position from that adopted by companies such as Klarna, which believed that AI agents could completely replace human capabilities.

Project management, due to its inherent complexity, even when has to deal with repetitive problems and activities, highlights a more dynamic (and probably also more ethically connoted) use of AI agents.

5.3 AI agents integrated into CRM: the case of Salesforce Einstein

Customer Relationship Management Systems are an essential component of business operations in today's economic environment and play a key role in maintaining customer relationships by ensuring optimal service quality and thus helping to increase revenues. Over time, CRMs have been subject to continuous innovation driven by increasingly pervasive technological innovation. The integration of AI agents into CRM has led to a further evolutionary leap in these systems, thanks to their ability to manage processes according to a data-driven logic, automation, and the increasingly significant personalization that AI agents are able to develop in their interaction with customers. This is referred to as "AI-powered CRM," a term that summarizes the interaction of approaches to customer relationship management in which AI agents play a fundamental role. However, it is worth noting that CRM has a dual defining meaning: even before referring to a set of integrated technologies and software for interacting with current and potential customers, it refers to a business strategy aimed at improving the company's commercial relationships, driving growth and development, and ultimately ensuring the company's sustainability in the medium to long term. In this second perspective, the concept of CRM broadens its scope to a dimension that goes beyond technology, involving the strategic and cultural dimensions that characterize every company. There are three main areas in which CRM performs its function:

- **Marketing:** CRM helps identify potential customers (*leads*), manage marketing campaigns, and monitor the effectiveness of promotional initiatives by tracking interactions and customizing communication processes.
- **Sales:** CRM provides support throughout the entire sales cycle, from *lead* qualification to opportunity management and contract closure. It enables the tracking of communications, proposals, and activities, providing a complete view of the customer relationship.
- **Customer Service:** Provides tools for managing support requests, complaints, and post-sales interactions. A good CRM allows operators to have a complete

historical view of the customer, improving the quality and speed of service (Kalaiyarasan, 2023).

AI can intervene in all these processes and has revolutionized the traditional concept of CRM, enriching it with predictive analytics capabilities, allowing us to move beyond the approach that sees CRM as "simple" data management systems to offer insights and automations that can improve process quality. The impact of AI on CRM is evident in several areas:

The impact of AI on CRM manifests itself in several areas:

- **Intelligent Automation:** AI automates repetitive, low-value tasks such as email categorization, automatic customer record updates, and follow-up scheduling. This frees sales and support staff to focus on more complex and strategic interactions.

- **Predictive Analytics:** AI analyzes large volumes of historical data to predict future behavior. For example, it can predict which *leads* are most likely to convert (*predictive lead scoring*), which customers are at risk of leaving the company (*churn prediction*), and which sales opportunities are most likely to succeed.

- **Large-scale personalization:** AI-powered recommendation engines can suggest personalized products or services to customers in real time, improving the shopping experience. In customer service, an AI agent can suggest relevant responses to agents based on the context of the conversation.

- **Conversational agents:** AI-based chatbots and virtual assistants can handle routine requests 24/7, providing immediate answers and routing more complex cases to human operators.

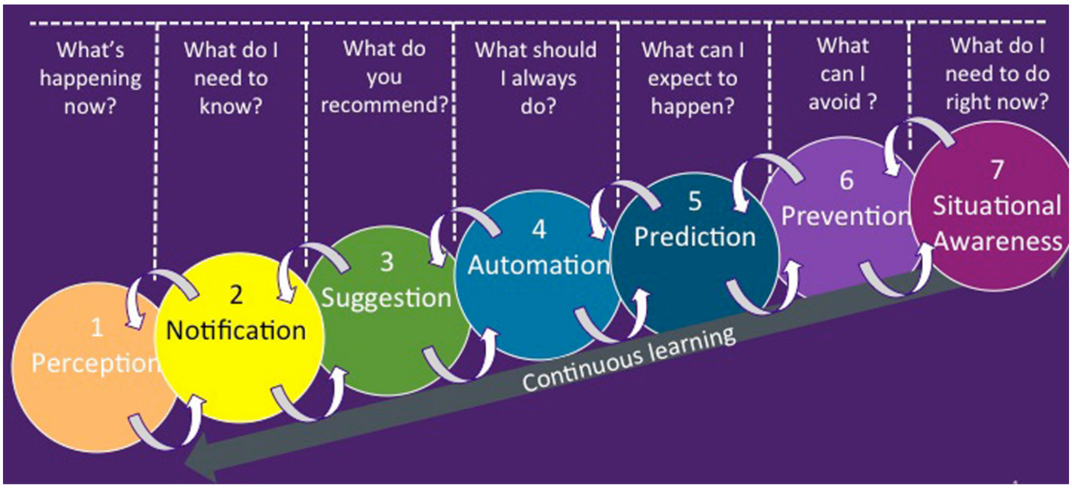
Salesforce Einstein is an interesting solution in this area. It operates on a global cloud infrastructure and promises to deliver high performance and security. It is based on a "multitenant" model, which allows companies to store their data in a separate and secure manner while sharing the power and specificities of the shared platform. This strategic decision is supported by the API-first design, which allows Salesforce Einstein to be easily integrated into other business systems, enabling economies of scale to access high-quality services while maintaining high security standards (Henschen, 2025).

The Salesforce developers chose not to offer an AI agent that attempts to mimic human behavior, but rather a system capable of improving customer relationship management, offering companies the ability to understand market trends, predict user behavior, and provide recommendations, automating only the most repetitive tasks but ensuring strong human-machine interaction. Salesforce Einstein fits into the principle of "augmented humanity" that we have already mentioned and which distinguishes it from other examples (such as Klarna). The goal of Salesforce Einstein is to integrate and assist human operators, not replace them, thanks to the potential of machine learning and natural language that allows users to interact directly with the AI agent.

Salesforce offers a wide range of solutions, all characterized by the concept of scalability. Einstein *Bot Builder* allows the creation, customization, and management of intelligent conversational agents without requiring in-depth programming skills: the API-first approach we mentioned earlier takes the form of an intuitive interface that does not require writing code, allowing administrators to quickly develop chatbots using predefined templates and reusable components. This approach significantly speeds up development times and facilitates the adoption of the technology even in business contexts with limited or no IT resources. The usability of the AI Agent is further enhanced by the *Skills Builder* and *Prompt Builder* applications, which allow organizations to define specific behaviors for chatbots, automate complex workflows, and ensure that virtual assistants' responses are consistent with the communication style and brand identity. Native integration with CRM data allows AI agents to access real-time information, improving the relevance of customer interactions and supporting data-driven decision- . Another key element is the *Model Builder:* a tool that develops and monitors complex scenarios and allows companies to develop a modular and scalable approach to introducing AI into customer relationship processes, enabling AI agents to customize their analytical and predictive capabilities in line with the needs of different areas of the organization. These tools enable substantial integration between conversational automation and core organizational processes, based on the use of NLP (Natural Language Processing) algorithms and machine learning, which allow chatbots to understand users' natural language, anticipate their needs, and provide accurate and efficient responses (Nursyafiqah, 2025).

From an architectural and algorithmic standpoint, Salesforce Einstein shares the advantages of AI agents' continuous learning cycles, with algorithms that, after the training phase, can operate through a seven-step process that helps improve system performance (Fig. 32):

Fig. 32 – Continuum learning model and Outcome of Salesforce Agent AI



Source: Henschen, 2025

To operate at its best, Salesforce Einstein needs a large amount of data to support its predictive capabilities based on machine learning: to offer adequate and consistent support, Salesforce needs at least 150 leads and 150 converted leads per month, and at least 10,000 lines are necessary for the statistical outputs to be reliable. These aspects are critical because, naturally, not all companies manage data volumes capable of supporting Salesforce's reliability. The main challenges are data and computing power. From this point of view, the problems are not only technological: the large amount of data available to the company is a competitive advantage in training algorithms and models, but it requires extra attention to protect data ownership and user privacy.

The specificity of the Salesforce Einstein business model is particularly interesting: the decision to offer customized solutions to companies without committing organizations to changes in their governance structure to host AI systems of a certain complexity is certainly a competitive advantage. however, as we have pointed out, the real crux of the system is linked to the need to work with large amounts of data, which are essential for building algorithms capable of making reliable predictions and models. From this point of view, Salesforce suffers from lower market penetration than other competitors such as Microsoft or IBM (Henschen, 2025).

5.4 The role of AI agents in UX through logistics optimization: the example of IKEA

The role of AI agents is fundamental to the user experience (UX). The concept of User Experience dates back to the mid-1990s and, in its original meaning, was much broader than the idea of human-machine interaction and usability that it has taken on in a highly technological context. Various attempts to systematize the definition have resulted in the International Standard on ergonomics of human-system interaction (ISO 9241-210), where UX is defined as

> "[...] a person's perceptions and responses that result from the use or anticipated use of a product, system or service."

This approach confirms that the user experience outlines a much broader conceptual framework than that relating to the usability of a graphical interface, to include a series of subjective aspects such as emotions, perceptions, satisfaction, and perceived values throughout the entire chain of interrelationships relating to the purchase, thus also including pre- nd post- s. The main goal of UX is to make interaction as intuitive, effective, and enjoyable as possible for the user, ensuring that the use of the product or service is not limited to its functionality, but solves a real problem and does so in a satisfactory manner (Berni et al, 2021).

The quality of UX is crucial to the success of a product because it can directly influence:

- Customer satisfaction and loyalty;
- Conversion rates and sales;
- Brand reputation;

This brief digression is particularly important in relation to the case study presented in this section, which concerns one of the leading companies in the furniture manufacturing and sales sector. IKEA has always linked its UX quite strongly to its overall logistics processes, to the point that it has been identified as one of the most innovative companies in the sector. The choice of packaging that is optimized and designed not only to facilitate the supply chain but also to be easily transported by the user, the management of an effective assembly system that can be handled independently by consumers even without specific DIY knowledge, thanks in part to the targeted design of the components, have made IKEA a true benchmark in the industry.

Ikea's digital transformation journey has seen the use of AI agents for various purposes, all aimed at improving the UX and in line with the intuition that has always guided the Swedish giant's business philosophy, which has always focused on supply chain efficiency and a shopping experience capable of adapting to a very wide range of customers in order to maintain its competitiveness in the markets and build customer loyalty.

IKEA's logistics are renowned for their scale and complexity, and the focus on particularly effective packaging, considered a strategic component capable of supporting the entire supply chain, has contributed significantly to an optimal UX and optimal supply chain (Hellstroem, 2011). The introduction of AI agents follows an overall strategy that, since at least 2017, has seen IKEA use technology to improve customer relations. In 2017, the IKEA Place app was launched, allowing customers to use an augmented reality tool to position IKEA furniture in their homes: viewing the furniture in a 3D environment has enabled the 8 million users who have downloaded the app to reduce purchase uncertainty. The app's accuracy in positioning furniture (estimated at 98%) has delivered excellent results, to the point that IKEA Place users are 11% more likely to convert their interest into a purchase than those who do not use the app. In addition, IKEA has also introduced self-service kiosks in its stores where users can browse the digital catalog, check stock availability, and identify products in stock.

Finally, from a commercial point of view, IKEA has developed a cross-selling strategy between its digital platform and physical stores, thanks to the "click-and-collect" model: customers can buy online and pick up the product in store. This strategy has yielded excellent results thanks to the potential for upselling and cross-selling. Picking up products in physical stores often leads to new purchases by customers: according to some estimates, users who buy online and pick up at physical stores spend 35% more than online users who choose home delivery.

The introduction of AI has therefore come at a time when the company was already ready to embrace this innovation, thanks in part to an ongoing training program for employees and collaborators. IKEA uses AI agents and machine learning for supply chain management and inventory management, ensuring product availability and keeping costs down. In particular, IKEA uses AI algorithms to predict demand and optimize inventory levels. The goal is to ensure that the most popular products are always available in the warehouse, while reducing the weight of surplus items that are less in demand. AI has improved operational efficiency and customer satisfaction (who no longer experience the frustration of not finding the product they want). According to some studies, the reduction in costs (15%) is associated with a significantly greater increase in service and customer satisfaction (30%). Furthermore, the use of AI and robotics to improve the efficiency and accuracy of warehouse management has an equally significant impact, estimated at around 40% of logistics costs, with significant benefits for the company and its customers (Patov, 2024).

In fact, IKEA's strategy regarding the adoption of AI agents is in line with what we might call a long-term approach that sees the deep integration of technology in all business processes, from design to supply chain management, from sales to customer relationship management. In this way, IKEA is able to offer its customers shopping experiences that live up to their expectations and has managed to face the challenges that have arisen since the pandemic. Digital transformation has involved the integration of digital technology in all areas of the business, including the company's external image (Yani, 2025).

5.5 Comparative analysis: efficiency and impact of AI agents on organizations and business performance

The analysis of the case studies presented in this chapter highlights some crucial aspects relating to the operational efficiency and impact of AI agents on organizations and their business performance. The application of these technologies clearly demonstrates tangible benefits, but also highlights challenges and limitations that companies face when deciding to adopt AI agents or other solutions based on artificial intelligence algorithms. Table 3 identifies, for each case study, the type of AI agent used, the efficiency metrics, the organizational impact, the effects on performance, and the main critical issues encountered:

Table 3 – Key elements of the case studies analyzed

CASE	TYPE OF AI AGENT	EFFICIENCY METRICS	ORGANIZATIONAL IMPACT	EFFECTS ON PERFORMANCE	CRITICAL ISSUES
Klarna	Autonomous conversational assistant	Reduction in request resolution time (82% less), management of 66% of chats.	Replacement of 700 FTEs (Full-Time Equivalents), but with subsequent rebalancing to integrate human support.	Estimated profit improvement of $40 million.	Decrease in customer relationship quality in the medium to long term
AutoGPT	Autonomous agent (self-prompting)	Acceleration of project cycles, reduction in data collection and analysis time.	Automation of complex tasks, freeing project managers for strategic activities.	Automation of complex tasks, freeing project managers for strategic activities. Indirect benefits in terms of cost optimization and time-to-market reduction.	Staff training and integration of different components in a collaborative environment
Salesforce Einstein	Predictive and collaborative agent integrated into CRM	Improved lead scoring, increased sales agent and support productivity.	Workforce enhancement, focus on value-added interactions, not replacement.	Increased conversion rates, improved customer loyalty, revenue growth (not directly quantified in the documents).	Availability of large amounts of data to enable accurate and statistically reliable analysis.
IKEA	Analytical (Logistics) and Conversational (UX) Agents	Supply chain optimization, reduction of logistics and inventory costs.	Operational efficiency, better resource management, integration of AI into marketing and sales processes.	Increased customer satisfaction and conversion rates (through improved UX) and reduced operating costs.	The model is effective if includes internal technology development planning based on scalability and continuous employee training.

Source: personal draft

From an operational efficiency standpoint, the introduction of AI agents has always represented a moment of growth and development in environmental performance: in all case studies, these improvements were evident in the organizational component, in customer relations, and ultimately in the company's financial performance. The analysis confirms what has emerged from the literature regarding the potential of AI agents to positively impact repetitive tasks, reduce response times in customer relations, and manage human and material resources more effectively in internal operational processes. However, the case studies have shown that in some cases the benefits are short-lived. Klarna's experience, for example, clearly shows that the initial adoption of AI agents can drastically reduce customer interaction times and improve response efficiency, but can even be counterproductive in the medium to long term. Similarly, AutoGPT shows how introducing AI into project management can increase decision-making accuracy and improve resource allocation, enabling greater control over project progress and creating a more efficient and data-driven organizational environment. The IKEA case shows that the potential for medium- to long-term benefits is greater in contexts that adopt a planned and scalable strategy for using technology, so that the entire corporate culture is ready to embrace it.

The comparison clearly shows that efficiency gains are not enough unless accompanied by an adequate organizational strategy that balances their use with human input. The Klarna case is emblematic in this sense, highlighting how an overly automation-focused strategy may initially produce positive results but lead to a deterioration in customer relations and, consequently, overall business performance in the long term. Over-reliance on AI agents in customer experience management risks causing organizations to lose the ability to manage relationships in an empathetic and personal way. Klarna, like many other companies that have been tempted by the possibility of replacing staff with AI, has had to take note of this approach and reassess its strategy.

From an organizational impact perspective, the integration of AI agents involves profound transformations that require careful management of change processes and have a particularly significant impact from a social perspective. The reduction in staff dedicated to low-value-added operational activities places the burden of retraining human resources on organizations. However, the best strategy seems to be that adopted by Salesforce Einstein, which represents a virtuous case: AI can be successfully used to enhance human capital rather than to replace it entirely. The 'augmented humanity' approach adopted by Salesforce emphasises the importance of using AI to enhance the human contribution to strategic activities and complex interactions, thus placing the complementarity between humans and machines at the centre of organisational action.

At the same time, the IKEA case highlights the potential of AI in continuously improving the user experience through the optimization of logistics and inventory management. IKEA demonstrates how the strategic and targeted application of AI agents, integrated with existing digital technologies, can generate benefits in terms of both cost reduction and increased customer satisfaction, thus providing a sustainable model in the long term. The analysis of IKEA's experience shows that technology must be introduced gradually: it shows consistent and long-term benefits if it becomes part of the corporate culture and is not perceived by staff as a risk, but as an opportunity. The IKEA example is particularly interesting because the adoption of AI Agents comes after a complex and planned technological innovation process, rather than being episodic.

The comparative analysis shows that AI agents can certainly improve business performance, but their success depends heavily on the organization's ability to manage their integration strategically. The key factor for the future development of AI is the ability to balance automation and human intervention, managing the organizational transition appropriately and ensuring consistency between the use of AI Agents and the company's strategic objectives. Only in this way will it be possible to maximize the benefits of AI while avoiding the risks and inefficiencies that would result from a superficial or ill-considered use of these technologies.

CONCLUSIONS

In this study, we have outlined the complex evolutionary path that has led to the introduction of AI (and recommendation agents in particular) in an increasingly wide range of production processes: from marketing to customer relationship management, from logistics to warehouse management, the implications of using AI are increasingly broad and almost all-encompassing. The analysis carried out in this study has highlighted how the evolution of Artificial Intelligence, culminating in the development of Digital Agents, is not a stage in a process of technological evolution, but a real paradigm shift that redefines the foundations of business and social dynamics. The study described an era of transformation in which data, processed by increasingly sophisticated algorithms, has become the primary strategic resource, enabling new business models and unprecedented personalization of the customer experience.

The introduction of AI agents has proven capable of optimizing processes, improving operational efficiency, and supporting complex decisions, reducing human bias and offering more intelligent management of human capital. While intelligent automation frees employees from repetitive tasks so they can focus on higher value-added activities, it also raises significant questions about the future of work and the skills needed to operate in a context of increasing human-machine interaction.

It is not a path without challenges. Recommendation agents are set to become not only innovative tools (crucial for creating a competitive advantage for brands that have been quickest to seize the opportunities they offer), but also a structural element of normal business management, thanks in part to their ability to integrate, manage, and process large amounts of data from different sources and to integrate with existing corporate systems (CRM, ERP, etc.). However, their operational "normalization" introduces ethical and legal problems: recommendation agents promise greater efficiency and cost reduction, but they also entail a profound change in industrial relations. Some operational processes that until now have been entirely managed by humans are being replaced (or integrated) by AI agents; roles within organizations are undergoing a process of change that seems not only irreversible but increasingly profound, to the point of suggesting that some professions (those characterized by a high degree of standardization and repetitiveness of processes) will be completely supplanted by AI agents. On the other hand, users will increasingly find themselves interacting with AI Agents when making choices online: whether it is finding content on major e-commerce platforms or choosing entertainment programs, suggestions from recommendation Agents are playing an increasingly decisive role. The algorithms that enable Agents to provide effective and balanced advice strike a balance between accuracy, which comes from including recommendations based on the user's previous purchasing experiences or interactions while browsing, and breadth, which allows the user to be presented with unfamiliar items, resulting in what is known in the literature as serendipity, or the pleasant feeling of discovering previously unknown options. From this point of view, recommendation algorithms are not objective in themselves; on the contrary, the capabilities of this technology also conceal a risk of potential manipulation, because suggestions are able to influence individuals on a large scale and in a more subtle, automated, and pervasive way than in the past. In fact, the literature refers to "digital manipulation" when influence is exerted through the use of technology; This can be intentional (when the algorithm is designed to deceive the individual's ability to choose according to their own needs and rationality), but it can also depend on factors beyond the control of developers and linked to users' limited awareness of how the recommendation agents they interact with work.

The practical application of AI agents, examined through case studies of Klarna, AutoGPT, Salesforce, and IKEA, has highlighted both the enormous potential benefits of using AI in production processes and the risks associated with the massive use of AI-based recommendation algorithms. The integration of these technologies requires particular attention to the management of organizational change and the relationships that are established between individuals both within the company and in the complex interaction between the needs and expectations of the brand and users. The analysis confirmed what emerged from the literature review: AI agents are powerful tools, but to fully exploit their potential, companies and th s will have to overcome technological fragmentation and integrate AI holistically, redefining information flows and promoting a corporate culture that values collaboration between humans and machines. This applies both from a human capital management perspective, because AI proves to be more effective when integrated with human talent. From this point of view, the cases of Klarna and IKEA show the two sides of the coin: on the one hand, the claim to use AI agents to replace labor, an attempt that, as we have seen, is only fruitful in the short term and forces companies to rethink their personnel strategies in order not to lose their reputation; on the other, the gradual integration of AI into production processes, which enhances a long-term trend of integration between technologies and professional skills within companies.

Our analysis has a number of implications for future research: these implications involve different disciplines and show that, despite enormous technological progress, the evolution of AI agents is not yet complete. In particular, the main areas in which research can help improve the impact of AI agents are:

- **Technological and architectural development:** research into technological architecture and the scalability of AI agents appears to be one of the most significant challenges. The need to invest in this particular area mainly concerns the ability of AI agents to manage complex tasks autonomously (multi-step) and their interaction with the environment. In particular, it appears necessary to improve the reliability and transparency of agents: results that require significant investment, also with a view to overcoming the technological and cultural barriers that may limit their use by end users.
- **Ethical and governance perspectives:** these are priority areas of research: the speed with which technological change is taking place requires the creation of a robust governance system capable of responding in a functional, rapid, and flexible manner to the ethical and regulatory challenges posed by the use of AI agents. It is essential to explore how to mitigate algorithmic bias, ensure data security and user privacy, and define a regulatory framework that is both effective and non- r innovation. Future research must be able to strike a balance between the growing autonomy of agents and human responsibility and supervision. It must be able to suggest regulatory and ethical responses quickly, offering businesses and citizens a reference framework that supports integration but prevents it from undermining the system of rights (social, labor, human).
- **Organizational and Social Impact:** Research must continue to investigate the impact of AI agents on the labor market, analyzing not only the phenomenon of "replacement anxiety," but also defining new professional and leadership opportunities. To do this, it is increasingly necessary to promote training programs

that offer all workers the skills they need to seize the new professional opportunities that will open up in the near future. It will be essential to study how organizations can implement effective *reskilling* and *upskilling* policies to accompany employees in this transition and how leadership must evolve to manage hybrid teams composed of humans and AI agents.

Recommendations

Based on the conclusions reached, the following practical recommendations are proposed for organizations and policymakers:

1. **Invest in a holistic AI strategy:** Companies should not implement AI in a fragmented, tactical manner (to gain a short-term competitive advantage) or solely to maximize profit. The potential of AI agents is enormous, but their use will be all the more effective if their introduction into production processes is the result of an integrated strategy that links the adoption of AI agents to business objectives, existing workflows, and ultimately to corporate culture. This requires the formation of cross-functional teams and investment in high-quality data infrastructure: companies that have invested over time in offering their employees a real "culture of innovation" appear to have an advantage, as demonstrated by the IKEA case. After all, influencing corporate culture requires time to allow the new approach to take root in the organizational structure and in people's behavior;

2. **Prioritize Ethics and Transparency:** The design and use of AI agents must be guided by sound ethical principles. Companies must be transparent about how data is used and how algorithms make decisions, especially in sensitive contexts such as HR or customer service. The creation of "AI ethics committees" can help monitor and guide responsible development.

3. **Promote Reskilling and Upskilling:** To mitigate replacement anxiety and capitalize on human potential, organizations must invest in continuous training programs. The goal is not only to teach the use of AI tools, but also to develop complementary skills, such as critical thinking, creativity, and emotional intelligence, which cannot be replicated by machines. The enhancement of human

capital continues to be a fundamental necessity for businesses if they want to maintain their competitiveness over time.

4. **Adopting an Agile Leadership Model:** We have seen how the main characteristic of tomorrow's leaders is the ability to facilitate process integration, while abandoning the traditional approach, which is oriented towards the mere formal control of behaviors, actions, and procedures. For businesses, this translates into the ability to recruit, train, and empower leaders who are capable of understanding the *insights* generated by AI and translating them into practical strategies, promoting an environment of continuous learning and adaptation.
5. **Collaborate on Regulation and Standards:** Governments and international organizations should work together to create a harmonized and flexible regulatory framework (such as the EU's AI Act) that protects citizens' rights without stifling innovation. Collaboration between the public and private sectors is essential to develop shared technical standards and ensure that AI agents operate safely and fairly.

BIBLIOGRAPHY

Abbracciavento, F., Formentin, S., Gualandi, Em Nanni, R., Paoli, A. Savaresi, S. M. (2020), "Modeling and prediction for optimal Human Resources Management," *IFAC-Papers OnLine*, 53 (2), pp. 16996-17001

AIleaders (2024), *Case Study: Klarna*, www.aileaders.eu

Alekseeva, N., Strganova, O., Vasilennok, V. (2019), "Identifying trends in the development of marketing in the digital age", *Atlantis Highlights in Computer Sciences*, 1, pp. 13-16.

Anderson, A. W., Li, T., Vorvoreanu, M., & Burnett, M. (2021), "Human-AI Interaction for Diverse Humans: What Cognitive Style Disaggregation Reveals", *arXiv*, Cornell University, doi.org/10.48550/arxiv.2108.00588

Andriopouls, C., Lewis, M. W. (2009), "Exploitation-Exploration tensions and organizational ambidesterity: managing paradoxes of innovation", *Organization Science*, 20 (4), pp. 696-717.

Barbera, D. (2025), "Candy Crush, 200 workers will be replaced by the AI they helped develop", *Wired*, July 16, 2025. https://www.wired.it/article/candy-crush-licenziamenti-ai/

Basiru, J. O., Ejiofor, C. L., Onukwulu, E. C. Attah, R. U. (2023), "Optimizing Administrative Operations: A Conceptual Framework for Strategic Resource Management in Corporate Settings", *International Journal of Multidisciplinary Research and Growth Evaluation*, 4 (1), pp. 760-773.

Berni, A., Borgianni, Y. (2021), "From the definition of user experience to a framework to classify its application in design", *International Conference on Engineering design*, Gothenburg, Sweden, 16-20 August 2021, pp. 1627-1636.

Bhagat, I. A., Wankhede, K. G., Kopwar, N. A. Sananse, D. A. (2024), "Artificial Intelligence in Healthcare: a Review", *International Journal of Scientific Research in Science, Engineering and Technology*, 11 (4), pp. 133-138.

Bîrzu, S. (2023), "Tracing the evolution of digital marketing from the 1980s", *Review of Management and Economic Engineering*, 22, 1 (87), pp. 73-77

Bokhari R. H. (2019), "ERP issues and challenges: a research synthesis", *Kibernetes*, 49 (3), pp. 629-659.

Bousetuane, F. (2025), "Agentic Systems: a guide to transforming industries with vertical AI Agents", Cornell University, pp. 3-31.

Briggs, J., Kodnani, D:, Hatzius, J., Pierdomenico, G., (2023), "The Potentially large effects of artificial intelligence on economic growth", Goldman Sachs Economics Research,

Bruno, Z. (2024), "The Impact of Artificial Intelligence on Business Operations", *Global Journal of Management and Business Research. D. Accounting and Auditing*, 24 (1), pp. 1-8

Brustoloni, J. C. (1991), "Autonomous Agents: Characterization and Requirements", *Technical Report* CMU-CS-91-204, School of Computer Science, Carnegie Mellon University.

Buczak, A. L., Guven, E. (2016), "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion detection", *IEEE Communications Surveys & Tutorials*, 18 (2), 1153-1176)

Buttle, F. A:, Maklan, S. (2015), Customer Relationship Management: concepts and technologies, pp. 3-23

Caffagni, D., Cocchi, F., Barsellotti, L., Moratelli, N. et al. (2024), "The Revolution of Multimodal Large Language Models: a Survey", *Findings of the Association for Computational Linguistics, ACL* 2024, pp. 13590–13618

Cath, C., Wachter, S., Mittelstadt, B, Taddeo, M., Floridi, L. (2018), "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach", *Science and engineering ethics*. 24. 10.1007/s11948-017-9901-7.

Chen, J., Geng, Y., Chen, Z., Pan, J. Z. et al. (2022), "Zero-shot and Few-shot learning with Knowledge Graphs: a comprehensive Survey", *Proceedings of the IEEE*, vol. 111, no. 6, IEEE, 2023, pp. 653–685.

Chen, Y, Ji, C. Cai, Y., Yan, T., Su, B. (2024), "Deep reinforcement learning in autonomous car path planning and control: a survey", Cornell University, arxiv.org/abs/2404.00340.

Cichocki, A., Kuleshov, A.P., (2021), "Future Trends for Human-AI Collaboration: A Comprehensive Taxonomy of AI/AGI Using Multiple Intelligences and Learning Styles", *Computational Intelligence and Neuroscience*, 8893795

Choudhary, N., Dawani, D., Darda, P. (2019), "Digital marketing: the 'jargon' of new era", *Journal of Emerging Technologies and Innovation Research*, 6 (12), pp. 19-30.

Christadopouloous, K., Mourad, S., Mataric, M. (2024), "Agent thinking fast and slow: a talker-reasoner architecture", Cornell University, arXiv:2410.08328v1.

Dash, R., McMurtrey, M., Rebman, C., Kar, U. K. (2019), “Application of artificial intelligence in automation of supply chain management”, *Journal of strategic innovation and sustainability*, 14 (3), pp. 43-53

Digital 2025. Global overview report, We are social & Meltwater. https://wearesocial.com/uk/blog/2025/02/digital-2025/ [May 15, 2025]

El Attaoui, A., Boukhafla, A., Rhouas, S., El Hami, N. (2023), “Custom application programming interface data extractor applied to the Klarna e-commerce dataset”, *Indonesian Journal of Electrical Engineering and Computer Science*, 31 (3), pp. 1624-1632.

Erwanda, Y. T., Doli, E. R. (2024), “Comparative study and traditional marketing strategies in increasing product sales”, *Smart international management journal*, 1 (3), pp. 14-20.

Faini, F. (2020), “The governance of artificial intelligence between ethics and law”, *Politeia*, XXXVI, 137, pp. 59-82

Feng, K. J. K., McDonald, D. W., Zhang, A. X. (2025), “Levels of autonomy for AI Agents”, *arXiv*, Cornell University, doi.org/10.48550/arxiv.2506-12469

Fenn, J., Raskino, M. (2008), *Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time*, Boston, MA, Harvard Business Press, 2008

Ferrer, X., van Nuenen, T., Such, J. M., Cot´ e, M., & Criado, N. (2021), “Bias and discrimination in AI: A Cross-disciplinary perspective”, *IEEE Technology and Society Magazine*, 40 (2), pp. 72–80.

Flores, P. L. (2008), *Leadership and Change Management*, Milan, Franco Angeli

Franklin, S., Graesser, A. (1997), “Is it an Agent, or just a Program?” *Journal of Autonomous Agents and Multi-Agent Systems*, 1(1), pp. 21–35.

Gallo, A. (2014), “The value of keeping the right customer,” *Harvard Business Review*, https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

Garg, P., Beeram, D. (2024), “Large Language Model-based autonomous Agents”, *International Journal of Computer Trends and Technology*, 72 (5), pp. 151-162.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B. et al. (2020), “Generative Adversarial Networks”, *Communications of the ACM*, 63 (11), pp. 139-144.

Gosmar, D., Dahl, D., Coin, E. (2024), “Conversational AI Multi-Agent Interoperability, Universal Open APIs for Agentic Natural Language Multimodal Communications”, Cornell University, 10.48550/arXiv.2407.19438.

Gupta, R., Gupta, S., Parikh, R., Gupta, D., Javaheri, A., & Shaktawat, J. S. (2025), “Personalized Artificial General Intelligence (AGI) via Neuroscience-Inspired Continuous Learning Systems”, Cornell University, doi.org/10.48550/ARXIV.2504.20109

Halbouni, A., Gunawan, T. S. et al. (2022), “Machine Learning and Deep Learning Approaches for CyberSecurity: A review”, *IEEE Access*, 10, 19572-29585.

Hamilton, M. (2018), “The biased algorithm: evidence of disparate impact on Hispanics”, *American Criminal Law Review*, 56, pp. 1553-1577

Han, L., Mubarak, A., Baimagambetov, A. Polatidis, N., Baker, T. (2025), “A Survey of Generative Categories and Techniques in Multimodal Large Language Models”, *Association for Computer Machinery ACM*, 1, 1, pp. 1-43

Haun, L. (2025), “Klarna Claimed AI was doing the work of 700 people. Now It’s rehiring”, *Reworked*, https://www.reworked.co/employee-experience/klarna-claimed-ai-was-doing-the-work-of-700-people-now-its-rehiring/

Hellstorem, D., Nilsson, F. (2011), “Logistics-driven packaging innovation: a case study at IKEA”, *International Journal of Retail & Distribution Management*, 39 (9), pp. 638-657

Henschen, D. (2025), *Inside Salesforce Einstein Artificial Intelligence. A Look at Salesforce Einstein Capabilities, Use Cases and Challenges*, Constellation Research.

Hlatshwayo, M. A. (2025), “The Integration of Artificial Intelligence (AI) Into Business Processes”, 10.5281/zenodo.10893971.

Huang, Y., Yihang C., Zhang, H. et al., (2025) “Deep Research Agents: A Systematic Examination And Roadmap”, Cornell University, 10.48550/arXiv.2506.18096.

Illescas-Manzano, M. D., Vicente Lopez, N., Gonzales, N. A., Rodriguez, C. C. (2021), “Implementation of Chtabot in Online commerce, and open innovation”, *Journal of Open Innovation*, 7 (125), 103390.

Jatobà, M., Santos, J. Gutierriz, I., Moscon, D. et al. (2019), “Evolution of Artificial Intelligence Research in Human Resources,” *Procedia Computer Science*, 164, pp. 137-142.

Jovanović, M., Campbell, M. (2025), "Self-Directing AI: The Road to Fully Autonomous AI Agents," *Computer*, 58 (2), pp. 71-77

Kaggwa, S., Eleogu T. F., Okonkwo F., Farayola O., A. Et al. (2024), "AI in Decision Making: Transforming Business Strategies," *International Journal of Research and Scientific Innovation*, 10 (12), pp. 423-444

Kalaiyarasan, B., Gurumoorthy, K., Kamalakannan, A. (2023), "AI-Driven Customer Relationship Management (CRM): A review of implementation strategies", *International Conference on Computing Paradigms*, pp. 33-38

Kordzadeh, N., Ghasemaghaei, M. (2021), "Algorithmic bias: review, synthesis, and future research directions", *European Journal of Information Systems*, DOI:10.1080/0960085X.2021.1927212

Korzynski, P., Mazurek, G., Altmann, A. Ejdys, Kazlauskaite, R. Et al. (2023), "Generative artificial intelligence as a new context for management theories: analysis of ChatGPT", Central European Management Journal, 31 (1), pp. 3-13

Kumar, V., Ashraf, A. R., Nadeem, W. (2024), "AI-powered marketing: What, where, and how?", *International Journal of Information Management*, 77, 102783

Kuzior, A., Sira, M., Borzek, P. (2023), "Effect of artificial intelligence on the economy", *Scientific papers of Silesian University*, 176, pp. 319-331

Lachure, J. S., Verma, G. K., Doriya, R. (2023), *Deep Learning: Theory, Architectures and Applications in Speech, Image and Language Processing*, 10.2174/97898150792101230101

Lakshmipriyanka, A., Hrihararao, M., Prasanna, M., Deepika, Y (2023), "A study on Artificial Intelligence in Marketing", *International Journal for Multidisciplinary Research*, 5 (3), pp. 1-12.

Latapie, H., Kilic, O., Thorisson, K. R., Wang, P. Hammer, P. (2022), "Neurosymbolic systems of perception and cognition: the role of attention", *Frontiers in Psychology*, 13, 806397

LeCun, Y., Gengio, Y., Hinton, G. (2015), "Deep Learning", Nature, 521, pp. 436-444

Liu, Y. (2022), "Effect of Digital marketing capabilities and Blockchain Technology on organizational performance and psychology", *Frontiers in Psychology*, 12, 805393

Locatelli, R., Pepe, G., Salis, F. (2022), *Artificial Intelligence and Credit Risk. The Use of alternative data and methods in internal credit rating*, Palgrave Macmillan

Madanchian, M. (2024), "From Recruitment to Retention: AI Tools for Human Resource decision-making", *Applied Science*, 14 (24), 11750

Maldonado-Canca, L., Cabrera-Sánchez, J.-P., Casado-Molina, A.-M., Bermúdez-González, G. (2024), "AI in Companies' Production Processes: What Do Their CEOs Think?", *Journal of Global Information Management*, 32(1), pp. 1-29.

Mantymaki, M., Mindinnen, M, Birkstedt, T., Viljanen, M. (2022), "Defining organizational AI governance", *AI and Ethics*, 2, pp. 603-609

Market Us (2024), *Global Artificial Intelligence market size*. https://market.us/report/artificial-intelligence-market/ [20.05.2025]

Martinez, A. (2025), "55% of companies regret AI Layoffs, HR leaders need to take note", *The HR Digest*, https://www.thehrdigest./55-of-companies-regret-ai-layoffs-hr-leaders-need-to-take-notecom/.

Mathew,

Meshram, K, Mishra, U, Rathnayake, U. (2025), "Application of artificial intelligence in agri-tech, environmental and biodiversity conservation", *Array*, 26, 100412

Migliorisi, A. (2025), "Amazon CEO writes to employees: AI will reduce our workforce," *Il Sole 24 Ore*, June 18, 2025. https://www.ilsole24ore.com/art/il-ceo-amazon-scrive-dipendenti-l-ia-ridurra-nostra-forza-lavoro-AHpxYZJB

Mishra, R. K., Reddy, G. Y. S. Pathak, H. (2021), "The Understanding of Deep Learning: A Comprehensive Review", *Mathematical Problems in Engineering*, 55448884.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016), "The Ethics of Algorithms: Mapping the Debate", AI & Society, 31(1), pp. 1-21

Najana, M., Bhattacharya, S., Kewalramani, C., Pandiya, D. K. (2024), "AI and Organizational Transformation: Navigating the Future", *International Journal of Global Innovations and Solutions (IJGIS).* DOI:10.21428/e90189c8.03fab010

Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann.

Nursyafiqah, A. (2022), "Building Trustworthy AI Chatbots with Salesforce Einstein and Copilot AI", *International Journal of Scientific Research & Engineering Trends*, 8 (5), pp. 1483-1488.

Olazarn M. (1993), "A Sociological History of the Neural Network Controversy", Advances in Computers, 37, pp. 335-425

Oliveira M. I. S., Loscio B. F. (2019), "What is a Data Ecosystem", *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in Data Age*.

Pascucci, F., Savelli, E., Gistri, G. (2023), "How digital technologies reshape marketing: evidence from a qualitative investigation", *Italian Journal of Marketing*, pp. 27-58.

Patov, A: (2024), "How IKEA Uses Technology to improve Customer Experience (CX) in retail", *Renascence*, https://www.renascence.io/journal/how-ikea-uses-technology-to-improve-customer-experience-cx-in-retail.

Prasanth, A., Vadakkan D: J., Surendran, P., Thomas, B. (2023), "Role of artificial intelligence and Business Decision making", *International Journal of Advanced Computer Science and Application*, 14 (6), pp. 965-969.

Putnam, L. L. (1986), "Contradiction and paradoxes in organizations", Thayer L. (ed.), *Organizational communication: Emerging* perspectives, Norwood, NJ, Ablex Publishing Corporation, Vol. I, pp. 151–167).

Putnam L. L. (2015), "Moving beyond "both/and" approaches: Alternative strategies for managing paradoxical tensions", *Presented at the annual European Group on Organizational Studies Colloquium, Athens*, July 2-4.

Qian, J., Chen, J. (2025), "Remaining Vigilant" while "enjoying Prosperity": how Artificial Intelligence Usage impacts Employees'innovative behavior and practive skill development", Behavioral sciences

Raisch, S., Krakowski, S. (2021), "Artificial intelligence and management: The automation augmentation paradox", *Academy of Management Review*, 46(1), pp. 192–210.

Rangone, N., Megale, L. (2025), "Risks without rights? The EU AI Act's approach to AI in law and rule-making", *European Journal of Risk Regulation*, pp. 1-16

Ravina-Ripoli, R., Dominguez, J. M., Montanes-Del Rio, M. A. (2019), "Happiness management en la época de la Industria 4.0", *Retos*, 9 (18), pp. 189-202

Ren, Y., Liu, Y., Ji, T., Xu, X. (2025), "AI Agents and Agentic AI-Navigating a Plethora of Concepts for Future Manufacturing", ArXiv, Cornell University, preprint arXiv:2507.01376v1

Rosemblatt, F., (1957), *The Perceptron. A perceiving and recognizing automaton*, Buffalo, Cornell Aeronautical Laboratory.

Rožman, M., Oreški, D., Tominc, P. (2022), "Integrating artificial intelligence into a talent management model to increase the work engagement and performance of enterprises", *Frontiers in Psychology*, 13, 1014434

Russell S., Norvig, P. (2020), *Artificial Intelligence: A Modern Approach*, Pearson ed.

Samuel A. L. (1959), "Some studies in machine learning using the game of checkers", *IBM Journal of research and development*, 3(2), pp. 210-229

Sapkota, R., Roumeliotis, K. J., Karkee, M. (2025), "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges", 10.48550/arXiv.2505.10468.

Schoning, H. (2018), "Industry 4.0", *IT Information technology*, 60 (3), pp. 121-123.

Smith, F. L. (2020), “Quantum Technology Hype and National Security”, *Security Dialogue*, 51(5), pp. 499-516.

Smith, W. K., Lewis, M. (2011), “Toward a theory of paradox: Adynamic equilibrium model of organizing.” *Academy of Management Review, 36*(2), pp. 381–403.

Sposato, M., Dittmar, E. C. (2025), “The AI-powered future of digital transformation: enhancing organizations and leadership development”, *Journal of Work-Applied management*, 17 (1), DOI 10.1108/JWAM-02-2025-0039

Stryker, C. (2025), “Types of AI Agents”, IBM.com, https://www.ibm.com/think/topics/ai-agent-types [25.06.2025]

Tadimarri, A., Jangoan, S., Sharma, K. K., Gurusamy, A. (2024), “AI-Powered marketing: Transforming Consumer Engagement and Brand Growth”, *International Journal For Multidisciplinary Research*, 6 (2), pp. 1-11

Tadjer, A., Hong, A., Bratvold, R. B. (2021), “Machine learning based decline curve analysis for short-term oil production forecast”, Energy Exploration & Exploitation, 39, pp. 1-23

Talukdar W. (2025), “Security in Agentic and Multiagent Systems”, *Analytics*. https://pubsonline.informs.org/do/10.1287/LYTX.2025.02.01/full/

Toffler, A. (1987), *The Third Wave*, Milan, Sperling&Kupfer

Thorbecke, C. (2019), “New York probing Apple Card for alleged gender discrimination after viral tweet”, *ABC News*, November 11, 2019. https://abcnews.go.com/US/york-probing-apple-card-alleged-gender-discrimination-viral/story?id=66910300

Vasawani, A., Shaweer, N., Parmar, N., Uszkoreit, J. et al. (2017), “Attention is all you need”, *31st Conference on Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, pp. 1-11

West, M., *An Introduction to Deep Learning*. https://www.bouvet.no/bouvet-deler/an-introduction-to-deep-learning. [26.6.2025]

Woodridge, M., Jennings, N. R. (1995), “Intelligent agents: Theory and Practice”, *The Knowledge Engineering Review*, 10 (2), pp. 115-152

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y. et al. (2023), “The Rise and Potential of Large Language Model Based Agents: A Survey”. *arXiv*, :2309.07864v3.

Yani, M. F., Muhdiantini, C., Firdaus, F. A., Muhammad, H. A. (2025), “IKEA: Enhancing customer experience through digital transformation”, *JIPI*, 10 (2), pp. 1171-1179.

Zaman, K. (2022), “Transformation of marketing decisions through artificial intelligence and digital marketing”, *Journal of Marketing strategies*, 4 (2), pp. 353-364.

Zhang, J., Li, W., Want, L. (2024), “Exploring the Governance of Artificial Intelligence Ethics: Current Issues and Challenges”, International Journal of Global Economics and Management, 4 (2), pp. 59-64