JEP

Firenze University Press
https://oaj.fupress.net/index.php/jep

OPEN ACCESS

Citation: Cencig, V., & D'Almo, M. (2025) From automation to agency: The paradigm of agentic AI across technology, society and the ontology of work. *Journal of Emerging Perspectives* 2: 87-96. doi: 10.36253/jep-20885

Received: April 30, 2025

Revised: September 30, 2025

Published: December 20, 2025

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

From automation to agency: The paradigm of agentic AI across technology, society and the ontology of work

Valerio Cencig[1]*, Mario D'Almo[2]

[1] *Head of Compliance Digital & Data Transformation at Intesa, Italy*
[2] *Head of Digital Strategy and Intelligence Compliance Area in Intesa Sanpaolo, Italy*
valerio.cencig@intesasanpaolo.com
*Corresponding author

Abstract. The evolution of artificial intelligence is redefining the relationship between humans and machines, shifting from traditional automation toward systems endowed with agency. Unlike conventional AI, which primarily supports or executes predefined tasks, agentic AI systems are capable of autonomous decision-making, proactive goal generation, learning from experience and coordinated action within complex environments. This shift represents not only a technological advancement but an ontological transformation, as machines increasingly operate as cognitive agents rather than passive tools. This article adopts an interdisciplinary perspective to examine the paradigm of agentic AI, tracing its evolution from earlier forms of automation and outlining its defining characteristics, architectures and application domains. It then analyses the implications for work and organisational structures, highlighting how agentic systems reconfigure roles, redistribute cognitive labour and enable new forms of human–machine co-agency. Finally, the paper addresses the ethical, legal and governance challenges raised by autonomous agents, arguing for responsible adoption frameworks that preserve human centrality, accountability and social justice in emerging socio-technical systems.

Keywords: agentic Artificial Intelligence, autonomous systems, human-machine co-agency, distributed cognition, multi-agent systems, ethical ai governance, socio-technical systems.

1. INTRODUCTION

The history of technology has been marked by progressive redefinitions of the relationship between humans and machines, each characterised by the gradual delegation of human functions to artificial devices. In the first phase, which began with the Industrial Revolution, machines replaced human muscular strength: mechanical looms, steam engines and hydraulic presses relieved humans from physical labour whilst keeping decision-making control firmly in the hands of the operator.

The second phase, the electronic automation of the twentieth century, introduced systems capable of performing predefined sequences of operations

through programming: industrial robots and automated assembly lines replaced not only physical force but also the repetitiveness of gesture, though they remained constrained by explicit and deterministic instructions.

The third phase, the digital automation of recent decades, broadly expanded data processing capacity: information systems process data, manage transactions and coordinate complex processes, yet continue to operate within rigidly predefined parameters, without the ability to learn or adapt independently.

Contemporary artificial intelligence introduces a significant qualitative shift from previous technological eras. For the first time, artificial systems are not restricted to executing tasks based on predetermined algorithms; they learn from data, discern intricate patterns, generalise acquired knowledge to new contexts and, notably, make autonomous decisions in situations not explicitly anticipated by their designers. This progression from task execution to decision-making represents an ontological transformation: the machine advances from being a mere instrument to exhibiting characteristics of a cognitive agent. Employing methodologies such as machine learning, deep neural networks and reinforcement learning, AI systems acquire capabilities traditionally attributed to biological intelligence, including contextual perception, inferential reasoning, abstract learning and strategic adaptation.

In this artificial intelligence revolution, agentic AI represents a further paradigmatic shift. Whilst conventional AI systems typically operate in reactive or consultative mode, addressing predefined queries, classifying data or generating outputs on request, artificial agents exhibit proactivity and operational autonomy. These agents continuously monitor their environments, generate sub-goals, devise action plans, execute tasks independently of ongoing supervision and refine behaviour based on experiential learning. This agency is characterised by decision-making autonomy, operational initiative and dynamic adaptation, leading to fundamental changes in production models in both manufacturing and services.

In material production, agentic AI enables adaptive factories where networks of agents coordinate machinery, manage supply chains, optimise logistical processes and anticipate failures through predictive maintenance: all in real time and with minimal human intervention. Multi-agent systems autonomously negotiate production priorities, reallocate resources in response to demand fluctuations and implement energy-optimisation strategies that adapt dynamically to environmental and economic constraints.

In the service sector, comparable transformations are evident: financial agents independently manage investment portfolios by analysing global market conditions and adjusting strategies almost instantaneously; diagnostic agents support medical professionals by processing complex medical records and recommending personalised treatment strategies; educational agents customise learning materials according to the individual cognitive rhythms of students; customer service agents engage in complex interactions, continuously adapting through the analysis of previous conversations.

This transition extends beyond technical and operational domains, influencing our understanding of action, autonomy, and responsibility, as well as redefining the meaning of human work. When machines do not simply execute but decide, when they do not only respond but anticipate, when they do not simply assist but collaborate, the boundary between tool and cognitive partner becomes porous. Accordingly, it is necessary to re-examine philosophical, legal and ethical frameworks established in eras when agency was an exclusive human attribute.

This article seeks to analyse this transformation through an interdisciplinary perspective. First, it explores the historical evolution from mechanical automation to artificial agency. Next, it examines the architectures, distinctive features and application domains of agentic systems. Then, it shows applied sectors and case studies. Finally, the analysis focuses on the transformation of work and underscores the importance of responsible governance that maintains the centrality of the human subject and preserves dignity, autonomy and social justice in the age of human–machine cognitive collaboration.

2. FROM THE PARADIGM OF AUTOMATION TO THE PARADIGM OF AGENCY

Agentic artificial intelligence introduces a substantial discontinuity in the relationship between human beings and technological systems. Unlike previous technologies, which operate exclusively as extensions of human capabilities under the control of an operator, agentic AI does not merely replace humans in the execution of predefined tasks; it actively participates in the decision-making process, exhibiting progressively increasing degrees of autonomy.

To understand this transformation, it is useful to conceptualise decision-making autonomy as a continuous spectrum.

At the first level of the autonomy spectrum there are artificial intelligence systems (which we might define as 'traditional machine learning') that operate in a purely

consultative mode, supporting human decision-making. A representative example is the advisory tools used by financial and investment consultants: the AI system analyses thousands of parameters: historical asset performance, market volatility, correlations amongst asset classes, macroeconomic indicators, sentiment analysis derived from financial news, sectoral growth projections, and generates contextualised, personalised recommendations. It suggests portfolio rebalancing, identifies investment opportunities or signals risks of excessive concentration in specific sectors and proposes hedging strategies to protect capital from market shocks.

However, each recommendation is reviewed by the human consultant, who retains full decision-making authority: the consultant evaluates the guidance, contextualises it with his knowledge of the client, integrates qualitative considerations beyond the system's reach and determines the appropriate course of action. The system performs no operation until explicitly instructed by the human operator. In this scenario, AI works as an expert adviser that augments human cognitive capabilities without supplanting final judgement. Nonetheless, important questions arise: to what extent do human consultants comprehend the rationale underlying algorithmic recommendations? Are they able to exercise truly independent critical judgement, or are they, in practice, inclined to follow recommendations they do not fully understand?

Moving along the autonomy spectrum, we encounter intermediate configurations in which AI assumes greater decision-making authority whilst remaining under human supervision. A significant example is the advanced autopilot systems used in commercial aviation. Modern autopilots (especially those trained with AI, in both civil and military contexts) do not merely maintain preset headings and altitudes; they actively manage critical phases of flight: take-off, navigation through congested air corridors, dynamic adaptation to changing weather conditions and landing.

During cruising, the system continuously makes operational decisions: it optimises speed and altitude to minimise fuel consumption, adjusts the route to avoid turbulence identified via radar and manages the aircraft's aerodynamic configuration (flaps, slats, spoilers) to maintain optimal efficiency. Human pilots monitor these processes but intervene actively only in exceptional circumstances: technical malfunctions, emergencies, extreme weather conditions; unexpected events requiring sophisticated contextual judgement. This configuration corresponds to what automation engineers call 'human-on-the-loop': AI operates autonomously within predefined parameters, while a human oversees the decision-making steps and can intervene or override at any moment, though their involvement is not essential to the process.

This arrangement gives rise to new challenges. For example, *automation complacency:* the tendency of human operators to develop over-reliance on automatic systems, thus reducing vigilance, is documented in numerous aviation incidents. When systems operate reliably for long periods, human operators may become overconfident, paying less attention to flight parameters and losing readiness to detect anomalies. *Deskilling*, the progressive loss of competence due to lack of practice, is another growing concern: pilots who fly mostly with active autopilot may experience deterioration in manual flying skills, precisely those crucial in emergencies requiring them to take direct control. These issues are already present, though less acutely, with deterministic autopilot systems; but they are likely to be exacerbated by new AI-trained systems.

At the opposite side of the spectrum are agentic AI systems operating in complete autonomy, making and executing decisions without direct human supervision. The most emblematic example is high-frequency algorithmic trading platforms based on artificial intelligence (AI High-Frequency Trading, AI-HFT), where artificial agents autonomously make decisions and execute the corresponding financial operations. A typical AI-HFT system monitors hundreds of global markets simultaneously, analysing continuous streams of data: real-time order flows, execution prices, trading volumes, bid–ask spreads, changes in volatility, financial news processed through natural language processing (NLP) and technical indicators computed over multiple time windows. When it identifies an opportunity, the algorithm autonomously executes the appropriate operations.

Here, direct human intervention is not only absent but structurally impossible: the information involved, and the temporal scales (measured in milliseconds) lie beyond human cognitive reach. Human traders perform a radically different role: they design high-level algorithmic strategies, define risk parameters (maximum exposure per position, daily loss limits, liquidity requirements) and monitor aggregate performances, adjusting strategies and parameters in response to structural market changes. In this configuration, even more extreme situations may arise in which AI systems develop emergent behaviours not foreseen by designers, for example: cooperative dynamics with other trading systems, potentially even in violation of market rules.

The transformation from 'instrument' to 'agent' thus raises profound questions that transcend purely technical or practical issues.

3. AGENTIC INTELLIGENCE: DEFINITIONS AND FUNDAMENTAL CHARACTERISTICS

To understand the concept of an artificial agent and its transformative implication, it is preliminarily necessary to identify what distinguishes it from a software system or a traditional machine-learning model.

The main properties that differentiate intelligent agents from traditional software programmes are autonomy, reactivity, proactivity and sociality.

Autonomy means that the agent can operate without direct human intervention, controlling its own actions and internal state. Whilst acting within a predefined context, an autonomous agent is not simply a programme that executes specific instructions, but a system that pursues objectives by adapting its behaviour to circumstances and selecting the most appropriate strategies considering the context.

Reactivity indicates that the agent perceives the environment in which it operates and responds promptly to changes occurring within it, also interacting with the environment itself.

Proactivity means that the agent does not merely react to contextual stimuli but displays goal-driven behaviour, taking initiative to achieve objectives.

Sociality indicates that the agent is equipped to interact with other actors, agents or humans, with different roles: as an agent performing a specific task, or as a 'controller' or 'orchestrator' in relation to other agents. This property is crucial because it enables the creation of cooperative multi-agent systems, where coordination is essential, capable of performing highly complex tasks.

Starting from these fundamental properties, agents may be classified along multiple dimensions. Each dimension represents one aspect of their architecture and behaviour.

A basic taxonomy distinguishes agents according to the complexity of their internal decision-making mechanisms. Here are some examples.

Simple Reactive Agents

Simple reactive agents operate through direct, deterministic condition–action rules – often not AI-driven. When a certain environmental condition is perceived, they perform a specific action. A classic example is the robotic vacuum cleaner: if it detects an obstacle, it changes direction; if it detects dirt beneath its sensors, it activates intensified suction; if the battery is low, it returns to its charging base.

Model-Based Agents

Model-based AI agents create and maintain internal representations of the state of the world, even when it is not directly observable, integrating current perceptions with prior knowledge to infer the complete state. An autonomous vehicle, for instance, maintains a model of the position of other vehicles, pedestrians and traffic signs even when temporarily outside its visual field. This model is continuously updated through sensors and tracking algorithms. Model-based agents can manage perceptual uncertainty and make informed decisions even when information is incomplete.

Goal-Driven Agents

Goal-driven agents incorporate representations of objectives that guide their behaviour. Rather than simply reacting, these agents evaluate individual actions based on how much they bring the system closer to the assigned goals. This requires predictive capability: the agent must be able to simulate the consequences of various alternative actions and select the one that maximises the chances of success.

A logistical planning agent tasked with delivering parcels, for example, formulates sub-goals (such as dividing parcels into batches with different deadlines), formulates and evaluates possible alternatives (delivery sequences), assesses consequences (total time, fuel consumption, likelihood of delays) and chooses the optimal plan. Goal-based agents are more flexible than reactive ones: if environmental conditions change, they can reformulate plans rather than being trapped in the previous plan.

Utility-Based Agents

Utility-based agents refine the goal-based approach by introducing utility functions that quantify the desirability of different world states, measuring the utility attainable. Whereas goal-based agents operate in a binary mode (goal achieved or not), utility-based agents recognise degrees of success.

Learning Agents

Learning agents incorporate learning mechanisms that allow them to improve performance through experience. Such agents operate with evolving knowledge and adapt to the encountered patterns.

Complex Agent Networks

Finally, there are complex networks of agents, functioning as genuine 'digital nervous systems', enabling high scalability of agent capabilities across extensive and diverse problems and situations. These networks are built according to rigorous design principles, including:

- task decomposition across cooperating yet independent agents;
- technological neutrality with respect to vendors, allowing integration of agents developed by different providers;
- embedded self-regulatory policies to prevent undesirable network behaviours.

Agent networks also possess the ability to adapt to new scenarios without requiring reprogramming, to reorganise autonomously if an agent fails (resilience) and to communicate in ways that make the system function as a single coherent entity.

What are the enabling technologies?

Enabling technologies range from reinforcement learning (including multi-agent RL) to symbolic planning, hybrid neural–symbolic systems and agents that exploit "*Large Language models* (LLMs)" to plan and orchestrate tools.

We may therefore conclude that agentic intelligence is not merely sophisticated automation but a new model of distributed cognition, in which the application of intelligence to processes, historically centred in humans, is distributed between humans and machines, making the boundaries between biological and artificial intelligence fluid and blurred.

Navigating these new paradigms does not simply mean managing technological evolution; it requires a theoretical rethinking of what constitutes intelligence, agency and cognition in an era characterised by deeply integrated socio-technical systems.

4. AGENTIC AI IN ACTION: APPLICATION SECTORS AND CASE STUDIES

The emergence of agentic systems in real-world applications is already transforming industries, services, and professional practices, generating both new opportunities and unforeseen challenges.

Traditional AI in finance, healthcare, industry and energy predicts or classifies but does not act: robo-advisors generate suggestions; credit-scoring systems classify creditworthiness; traditional anti-fraud systems issue alerts; imaging diagnostics produce predictions; sensor systems signal the need for maintenance interventions, and so on.

Agentic AI changes the game because it produces operational outcomes.

We have already discussed how, in trading, agents maintain a representation of market state, plan sequences of actions and execute them: orders, hedging, strategy shifts as markets evolve.

In healthcare, hospital agents could continuously monitor critical patients, aggregate signals from multiple devices, detect deterioration, activate protocols, notify medical staff and document everything. In oncology, they may adjust dosages, integrate new evidence and compare treatments using digital twins (dynamic digital replicas). In the operating room, they may coordinate scheduling, personnel and equipment, reorganising the entire day if an emergency case arrives. In emergency departments, they may manage triage and queues in real time; in large-scale disasters, networks of agents may coordinate ambulances, hospitals, logistics and supplies, distributing patients and resources where most needed.

In industry and logistics, fleets of warehouse robots already negotiate tasks, optimise routes, avoid collisions and replan when the environment changes. In supply-chain management, agents monitor demand, rebalance inventories, negotiate with suppliers, orchestrate transport and autonomously react to unexpected events. In maintenance, they do more than issue predictions: they book downtime windows, order spare parts and update documentation upon completion.

Marketing will also need to evolve. The advent of autonomous AI agents could change how markets interact with customers because customers may no longer be human beings but bots and agents. Colours, sounds, advertising, essentially the entire symbolic toolkit of persuasion, could become irrelevant.

In the energy sector, domestic and industrial agents control solar panels, batteries, flexible loads and electric vehicles to optimise costs and comfort, whilst aggregators and grid-level agents balance supply and demand like a virtual power plant, intervening autonomously on peaks and shortages.

A highly advanced anti-money-laundering agent ('Financial Crime Agent') could operate 24/7 on transactions, accounts and digital channels; it would maintain a dynamic graph of relationships between customers, merchants, crypto wallets, companies and high-risk jurisdictions; and detect dynamic patterns identifying money-laundering activities. In the case of anomalies, and following a risk-based approach, it could:

1. block suspicious transactions and freeze connected accounts within authorised limits;
2. orchestrate enhanced due diligence, requesting documents from the customer and cross-checking them with third-party sources (company registries, sanctions lists, *Politically Exposed Persons* registry, previous reports);
3. perform cross-checks, asking a 'merchant agent' for evidence of the transaction (invoices, order confirmations), correlating timestamps and IP addresses and comparing them with the customer's historical behavioural patterns;
4. compile a Suspicious Activity Report for the regulator, automatically generating a clear, fact-based narrative supported by data and graphs;
5. update its internal rules based on feedback, reducing future false positives.

In short, traditional AI observes and predicts, whilst agentic AI observes, decides, acts, learns, improves and coordinates.

5. AGENTIC AI AND THE TRANSFORMATION OF WORK: WHEN MACHINES BECOME COLLABORATORS

Agentic artificial intelligence is reshaping the labour market in ways that diverge fundamentally from previous waves of automation. These are not merely tools that increase human productivity: artificial agents can participate actively in decision-making processes, learn from experience and autonomously adapt their strategies. This transformation does not replace human beings, but redefines their role, shifting focus from execution to strategic supervision, from direct control to the design of intelligent systems.

A stand-alone generative AI system, such as a Large Language Model, responds to requests, generates content on command and supports human cognitive activities, but remains fundamentally reactive. An artificial agent, on the other hand, can participate actively and proactively in the processes and decisions required.

This capacity for active participation in decision-making, not merely information support, distinguishes agentic AI and defines its specific impact on the labour market. The implications are significant:

- redefinition of operational responsibilities: human workers no longer supervise the execution of tasks but orchestrate ecosystems of agents operating with varying degrees of autonomy;
- increased decision-making speed: agents operate on time scales inaccessible to human cognition (milliseconds in algorithmic trading, micro-level real-time management in smart factories);
- emergence of new forms of collective intelligence: hybrid human–agent teams in which decisions emerge from cooperative interaction rather than traditional hierarchy.

New Role of Human Intelligence

Contrary to dystopian narratives predicting the full replacement of human labour, agentic AI repositions the human being at the centre of the system, but in a qualitatively different role.

This shift materialises in several ways:

- Definition of Objectives and Constraints. Agents operate within decision-making frameworks established by humans. An algorithmic trading agent may autonomously execute millions of transactions, but risk parameters, return targets and ethical constraints (e.g. exclusion of controversial sectors) are defined by human portfolio managers. This initial configuration requires advanced skills: deep domain expertise, the ability to formalise values into computational metrics and to anticipate unintended consequences;
- Continuous Supervision and Intervention by Exception. Even when agents operate autonomously, humans retain veto power and ultimate authority when necessary. Agentic systems include control mechanisms: abnormal situations, high-risk decisions or unforeseen scenarios trigger alerts that, under certain conditions, require human intervention;
- Handling Ethical Grey Areas. Many work situations involve ethical dilemmas without algorithmically determinable solutions. A fully autonomous 'Level 5' vehicle could, in principle, operate without supervision, but decisions concerning whom to protect in an unavoidable crash raise moral questions beyond mathematical optimisation. Programming agents inevitably reflect value-based choices: maximise passenger safety? Minimise total casualties? Prioritise younger lives? These decisions require human collective deliberation and cannot be delegated to autonomous systems;
- Legal Responsibility and Accountability. When an agent causes harm, a drone injures someone or an AI system discriminates unfairly, responsibility falls on human and institutional actors: designers, implementers, supervisors, organisations and firms. Legal systems do not grant legal personality to AI. This regulatory constraint ensures that humans remain accountable, incentivising robust, transparent and auditable system design. However, the distribution of responsibilities amongst multiple actors still needs to be further explored and clarified.

New Professions

The rise of agentic AI is generating professional categories specifically oriented towards the management, supervision and orchestration of artificial agents. These roles differ broadly from both traditional IT occupations and the newer roles associated with generative AI.

Examples include:

- Agentic Systems Architect. Designs architectures for multi-agent systems in which dozens or hundreds of agents cooperate, negotiate resources and coordinate autonomously. This role defines communication protocols, conflict-resolution mechanisms, resource allocation strategies and desirable emergent properties. Unlike a traditional software architect, they must anticipate unintended behaviours that may arise from interactions amongst autonomous agents.

- Agent Alignment Specialist. Ensures that the operational objectives of agents remain aligned with the organisation's and society's values. They use techniques such as Inverse Reinforcement Learning (which infers the 'reward' function underlying an agent's behaviour), implement Constitutional AI mechanisms to encode ethical principles and conduct systematic red teaming to identify unintended behaviours before deployment. This competence is crucial because continuously learning agents can drift progressively away from initial values (a phenomenon known as *value drift*).
- Multi-Agent Coordination Manager. Supervises fleets of autonomous agents operating in physical environments: drones in logistics, collaborative robots in manufacturing, autonomous vehicles in urban mobility. They do not directly control each agent but orchestrate collective behaviours, define dynamic priorities, intervene in anomalous situations and ensure safety. Unlike a traditional logistics manager, this role deals with agents making autonomous tactical decisions; supervision concerns meta-strategies and systemic risk prevention.
- Autonomous Systems Ethicist. Translates ethical principles into operational constraints for autonomous agents. When an organisation states that it wishes to operate 'responsibly' or 'fairly', this professional formalises these values into measurable metrics, verifiable tests and enforcement mechanisms. In finance, they define ethically acceptable trading strategies; in healthcare, they set fairness criteria for algorithms allocating scarce resources; in the justice system, they identify algorithmic biases violating principles of non-discrimination.
- Agentic Interaction Designer. Designs the interface through which humans communicate with agents and supervise their activities. Unlike a traditional UX designer, who optimises interactions with passive software, this role must account for agents that exhibit proactive behaviour, autonomous initiatives, and potentially unpredictable situations; for example, it requires designing ways to enable timely human intervention without demanding constant supervision.

New markets will also emerge, such as an insurance market specifically developed to transfer risks associated with agentic AI.

These professions require deeply interdisciplinary expertise:

- Technical skills in machine learning, multi-agent systems, game theory;
- Strong sectoral domain expertise (finance, healthcare, logistics, etc.) to anticipate practical implications of design choices;
- Ethical and philosophical skills to navigate value dilemmas, balance trade-offs and formalise moral principles into operational constraints;
- Communication skills to mediate between technical and non-technical stakeholders and translate business needs into technical specifications;
- Metacognition, the ability to reflect on one's cognitive processes and how they interact with those of artificial agents.

In addition to emerging professions, all existing professions, starting with cybersecurity, will need to evolve their paradigms within this new technological reality.

A large-scale programme of training and reskilling will be necessary to avoid widening the digital divide. This challenge must be addressed by both enterprises and individuals, with a lifelong learning perspective.

New Divisions of Labour: Co-Agency

The concept of *co-agency*, the cognitive collaboration between human and artificial intelligence, is the emerging paradigm in the labour market transformed by agentic AI. Unlike traditional automation, which replaced human tasks with mechanical processes, and unlike generative AI, which assists humans as an advanced support tool, agentic AI establishes operational partnerships where decisions and actions emerge from a coordinated interaction.

Dynamic division of cognitive labour

In a hybrid human–agent team, task allocation is not fixed but is reconfigured according to the context. In the near future, an operating room equipped with agentic robotic systems might be managed by a human surgeon making high-level strategic decisions (such as surgical approach), whilst robotic agents autonomously execute micrometric precision movements, compensate for tremors and optimise angles. If the system detects a tissue anomaly, the agent will call for the surgeon's intervention. Leadership flows fluidly between human and machine depending on the competencies required.

Structured cognitive complementarity

Humans and artificial agents possess complementary cognitive capacities that, when strategically integrated, exceed the performance of either working alone.

Agents excel at:

- processing massive volumes of data;
- detecting complex statistical patterns;
- multi-parameter optimisation;
- high-speed operations;

– consistent decision-making.

Humans maintain superiority in:

– contextual understanding and common sense;
– deep causal reasoning;
– creativity and lateral thinking;
– navigating ambiguity;
– ethical judgement in complex situations.

Agentic AI allows a systematic integration of these complementarities, not as a sum of separate contributions, but as emergent collective intelligence.

Reciprocal learning and continuous adaptation

In co-agency, not only do agents learn from experience, but humans progressively adapt their cognitive strategies in response to agent behaviour. A trader working with agentic high-frequency algorithms develops intuitions about market patterns initially recognised only by the agents. A radiologist collaborating with diagnostic agents refines attention towards subtle anomalies flagged by agents with high confidence. This co-evolutionary learning enhances and extends human capabilities.

Negotiation and conflict resolution

When networks of autonomous agents collaborate, conflicts inevitably arise from divergent goals to differing interpretations of the same situation and competing priorities. Advanced agentic systems incorporate negotiation and mediation mechanisms inspired by game theory and mechanism design theory.

In a multi-agent smart-grid energy-management system, agents representing producers, consumers and grid operators negotiate prices and allocations continuously. Humans do not manage these negotiations directly but define the meta-rules: what constraints apply? How should efficiency and fairness be balanced? What service levels are required?

This evolution will be accompanied by a shift from highly specialised, 'vertical' work roles towards more integrated roles across organisational departments, dissolving existing internal boundaries (siloes).

New Forms of Organisation

The adoption of agentic AI implies deep organisational transformations beyond technological implementation. Distributed, autonomous systems surpass traditional hierarchical structures, characterised by vertical command chains and centralised decision-making processes.

In organisations evolving towards agentic ecosystems:

– Control becomes coordination: managers do not make every decision but orchestrate systems of agents operating with delegated autonomy within defined parameters;
– Decision-making becomes distributed: there isn't a single decision-maker, but a network of human and artificial actors contributing complementary perspectives and tasks;
– Adaptation replaces rigid planning: instead of following rigid predefined plans, organisations set strategic objectives and allow agents to develop optimal tactics dynamically;
– Transparency becomes an operational necessity: to effectively coordinate autonomous agents, organisations must explicitly formalise objectives, constraints and success criteria that were previously implicit in corporate culture.

New psychological, social and human challenges

The deepest impact of agentic AI on the labour market will be psychological rather than merely economic. For centuries, personal and social identity has been intimately tied to work: what you do defines who you are. As artificial agents assume increasing portions of productive activity, the value of human contribution shifts from efficient execution to the capacity to orient, interpret and attribute meaning.

This transition raises profound philosophical questions:

– What is the role of the human being when operational efficiency is automatable? With agentic AI, humans define what is worth doing, not how to do it;
– How can a sense of fulfilment be maintained if tasks once considered rewarding are delegated to agents? Humans will need to recognise value in strategic supervision, complex orchestration and ethical mediation: domains that are inherently human;
– How should time be distributed amongst work, learning and leisure in an economy where productivity is maximised by agents? Agentic AI could enable shorter working hours without productivity loss, freeing time for creative, relational and caregiving activities – quintessentially human dimensions.

Hannah Arendt distinguishes between *labour* (activities necessary for biological survival), *work* (production of durable objects that transform the world) and *action* (interactions amongst humans that generate political and cultural meaning).

Agentic AI progressively assumes labour and portions of work, but leaves, and indeed expands, the space of action, reserved for humans.

Far from marginalising the human being, this evolution can reposition human intelligence at the centre, at a higher level of abstraction. Human beings become:

- Definers of goals and values: establishing what agents must pursue and what ethical constraints they must respect;
- Strategic supervisors: monitoring emergent behaviours, intervening in exceptional situations and ensuring continuous alignment;
- Interpreters of meaning: translating algorithmic outputs into contextual understanding and assessing implications beyond quantitative metrics;
- Ethical mediators: navigating moral dilemmas, balancing value trade-offs and assuming responsibility for automated decisions;
- Architects of 'cognitive symbiosis': designing ecosystems in which human and artificial intelligence collaborate effectively, maximising complementarities.

To avoid technological unemployment, the labour market in the era of agentic AI requires a radical reconfiguration of skills. Emerging professions (Agentic Systems Architect, Agent Alignment Specialist, Multi-Agent Coordination Manager, Autonomous Systems Ethicist) represent only the beginning. Practically every professional role will need to integrate competencies in supervising artificial agents, orchestrating autonomous systems and interpreting algorithmic decisions.

This transition requires substantial investment in training and reskilling but, above all, a cultural shift: from viewing AI as a threat to recognising it as an amplifier of human abilities; from competing with agents to collaborating with them; from preserving traditional tasks to imagining new ways to generate value.

The ultimate challenge is not (only) technical but human and existential: to build a cognitive symbiosis in which humans remain at the centre not despite agentic AI, but *thanks* to it. Artificial agents free human intelligence from repetitive, computationally intensive tasks, allowing concentration on what is intrinsically and uniquely human: creativity, empathy, ethical judgement, attribution of meaning and building meaningful relationships.

In this perspective, agentic AI does not diminish but amplifies human centrality, elevating humans from executors to strategic guides of technological and social evolution and, above all, of themselves.

6. TOWARDS RESPONSIBLE ADOPTION: GOVERNANCE, REGULATION AND FUTURE PERSPECTIVES

The integration of agentic intelligence into productive and social systems requires a clear and articulated ethical and political vision, embodied in fundamental principles, a regulatory framework, responsible practices at individual, organisational and systemic levels.

Europe, through the AI Act, seeks to position itself as a leader in trustworthy and responsible AI. This approach is not without criticism: some argue that excessive regulation undermines competitiveness; others argue that it is still insufficient to address systemic risks. The debate continues, reflecting genuine tensions amongst legitimate viewpoints.

The European AI Act, which has come into force progressively from 2024, adopts a risk-based approach:

- unacceptable practices (such as governmental social scoring or subliminal manipulation) are prohibited;
- high-risk systems (such as those used for credit assessment or the management of critical infrastructure) must meet stringent requirements concerning data quality, documentation, transparency, human supervision and robustness;
- other systems are not subject to mandatory requirements (except for certain transparency obligations).

The regulatory challenge is to balance risk protection with the promotion of innovation in a domain characterised by extremely dynamic technological evolution, in which generative and agentic models have joined traditional machine-learning approaches.

Excessively strict regulation could stifle research and development, whilst self-regulation would leave room for abuses and harm. Moreover, the global nature of AI technologies requires international coordination: fragmented standards could create trade barriers and opportunities for regulatory arbitrage.

Ensuring alignment, robustness, scalability, and interoperability requires continuous advances in research and development. However, the most profound challenges are moral and ethical. How should institutions and labour markets be redesigned for productive human–AI collaboration? How can economic benefits be distributed fairly, avoiding concentrations of power? How can systems that are natively opaque be developed to respect democratic values of transparency, accountability and fairness? How can human dignity and autonomy be preserved in cognitive ecosystems increasingly mediated by algorithms?

In these complex 'equations', the focus must remain not only on the human being but also on the planet as a whole, with an eco-centric vision, since humanity is not separate from, but deeply interdependent with the ecosystem.

The debate should not be driven by Big Tech alone; instead, it requires interdisciplinary dialogue. Technologists must engage with philosophers, lawyers with social scientists, policymakers with citizens, co-creating

shared visions of desirable futures and implementing regulatory frameworks that orient innovation towards the common good.

The full potential and social impact of agentic intelligence will depend on our collective ability to develop broad ethical maturity, one that can sustain new ways of creating meaning and assuming responsibility as social and technological complexity grows. Developing sophisticated algorithms is not enough: society must also cultivate collective wisdom in order to use them responsibly.

Agentic intelligence is not an inevitable fate that befalls us, but a social and technological construction that we can actively shape through individual, organisational and political choices. The crucial question is not whether intelligent machines will arrive (they are already here) or what they can do (increasingly more), but rather what kind of relationship we want with them and what kind of society we wish to build through, and with, them.

The challenge ahead is to build a cognitive symbiosis in which human and artificial intelligences are complementary, amplifying human capabilities without sacrificing values. This requires technical humility (acknowledging limits and uncertainties), ethical ambition (aiming for high standards of responsibility) and collective courage (facing difficult trade-offs with transparency).

Only through such a shared commitment can we ensure that the development of artificial intelligence will contribute positively to social, cultural and economic progress, serving humanity rather than exploiting it, enriching human experience rather than impoverishing it, expanding opportunities for self-realisation rather than generating new forms of alienation. The future of AI is not written. It is up to us, collectively, to determine it through conscious action, democratic deliberation and ethically grounded choices. The challenge is immense, but so is the opportunity to shape a better future for ourselves and for the generations that follow.