AI & Partners

Amsterdam - London - Singapore

EU AI Act

Fostering Agentic AI

How the development and functionality of AI agents – and the implications of their use – are fostered by the EU AI Act, particularly from a risk classification standpoint, amid rapid advances in large language and multimodal models.

March 2025

AI & Partners

AI & Partners

Amsterdam - London - Singapore

AI & Partners defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots professional services, regulatory interventions, and participating in industry groups such as AI Commons, we fight for fundamental rights in the artificial intelligence age.

This report was prepared by Sean Donald John Musch and Michael Charles Borrelli. For more information visit https://www.ai-and-partners.com/.

Contact: Michael Charles Borrelli | Director | m.borrelli@ai-and-partners.com.

This report is an AI & Partners publication.

Our analysis shows that the EU AI Act, as a legal framework for trustworthy AI, provides helps mitigate risks linked to agentic AI systems. These systems pose specific risks as they exhibit autonomy, goal-oriented behaviour, and environment interaction, affecting decisions made independently and strategies based on inputs. Specifically, mandating transparency and accountability addresses risks associated with their learning capabilities, ensuring ethical deployment. Also, governance practices enhance workflow optimization and multi-agent collaboration, supporting seamless system integration while mitigating risks. These measures establish a foundation for innovation, enabling agentic AI to transform business processes while maintaining trust within a digital ecosystem.

About this report

This report is based on market research, publicly available data, and interviews with AI specialists in AI & Partners, financial services organisations, and relevant third-parties. Moreover, quotations provided on specific topics reflect those of AI specialists at AI & Partners to be as representative as possible of real-world conditions. All references to EU AI Act reflect the version of text valid as at 13 June 2024. Accessible here.

Contents

Executive Summary

The development of agentic artificial intelligence (AI), characterized by its autonomous decision-making capabilities, is seen as a catalyst for profound transformations across industries, with first implementations already in place. Amid the accelerating advancements in large language models and multimodal systems that facilitated the rise to agentic AI, the European Union (EU) has introduced the AI Act, a groundbreaking regulatory framework designed to govern AI's development and deployment. Central to the Act is its risk classification approach, which categorizes AI applications based on their potential societal impact. The overall framework and the risk classification approach can help harness the potential of agentic AI in a robust and safe, yet effective way. While the risk-based regulation is of general nature, agnostic to the chosen technology, it serves as a baseline for the expectations on controls and governance in place when utilizing agentic AI approaches, such as transparency and human oversight obligations for high-risk cases.

This whitepaper examines the interplay between the EU AI Act and the evolving landscape of agentic AI, with a focus on regulatory, ethical, business, and societal implications.

Context and Purpose

The EU AI Act aims to establish a balanced regulatory environment that encourages innovation while mitigating risks. Its risk classification framework defines AI applications across four tiers: unacceptable risk, high risk, limited risk, and minimal risk. By providing a structured approach to governance, the Act seeks to ensure the safe and ethical deployment of AI systems while fostering public trust.

Agentic AI—systems capable of performing autonomous actions without direct human intervention—presents unique challenges and opportunities within this framework. These systems hold immense potential to improve efficiency and decision-making but also pose significant risks, including ethical concerns, data privacy challenges, and accountability issues. This whitepaper explores how the EU AI Act addresses these complexities and supports responsible AI innovation, in the context of Agentic AI.

Key Themes

1. Risk-Based Regulation The EU AI Act's tiered classification system underpins its regulatory strategy. High-risk applications, such as AI in critical infrastructure, healthcare, or recruitment, are subject to stringent requirements, including transparency, robustness, and bias mitigation. Meanwhile, prohibited applications—such as social scoring or manipulative systems—reflect the EU's commitment to upholding fundamental rights. While the risk-based regulation is of general nature, agnostic to the chosen technology, it serves as a baseline for the expectations on controls and governance in place when utilizing agentic AI approaches, such as transparency and human oversight obligations for high-risk cases.

2. Ethical and Societal Considerations The Act emphasizes the importance of human oversight and algorithmic transparency, ensuring that agentic AI operates within ethical boundaries. Provisions for user awareness and accountability mechanisms aim to address societal concerns, including fairness, equity, and inclusivity. The Act also promotes accessibility by mandating that AI systems accommodate diverse user needs.

3. **Business Impacts** For businesses, the EU AI Act represents both a challenge and an opportunity. Compliance with high-risk requirements may involve significant investments in documentation, testing, and auditing. However, the benefits—enhanced trust, competitive advantage, and alignment with global regulatory trends—position compliant businesses for long-term success. Regulatory sandboxes provide a controlled environment for innovation, allowing organizations to experiment with new technologies while adhering to safety standards. The development of agentic artificial intelligence (AI), characterized by its autonomous decision-making capabilities, is seen as a catalyst for profound transformations across industries, with first implementations already in place. Amid the accelerating advancements in large language models and multimodal systems that facilitated the rise to agentic AI, the European Union (EU) has introduced the AI Act, a groundbreaking regulatory framework designed to govern AI's development and deployment. Central to the Act is its risk classification approach, which categorizes AI applications based on their potential societal impact. These can be particularly useful for the use of agentic AI, in assessing the robustness when facilitating complex interactions between systems or between systems and humans.

4. **Global Influence** The EU AI Act's comprehensive approach has far-reaching implications beyond Europe. By setting a global benchmark for AI regulation, the Act encourages harmonization and raises the bar for ethical AI practices worldwide. Multinational companies must navigate this framework to ensure compliance, contributing to the global discourse on AI governance. These obligations aim for responsible development practices, which are particularly relevant for complex systems like agentic AI, fostering sustainable and safe use of such technology not only in Europe, but on global scale.

Introduction

The rapid evolution of artificial intelligence AI has reshaped numerous industries, driving innovation while introducing complex challenges. Among the transformative technologies, AI agents[1]—autonomous systems capable of perceiving their environment, making decisions, and acting to achieve goals—stand out as pivotal drivers of efficiency and innovation.

Their advanced functionality, particularly with breakthroughs in large language models (LLMs) and multimodal AI systems, offers unprecedented opportunities but also presents significant risks. Recognizing this dual nature, the European Union's AI Act establishes a comprehensive regulatory framework to govern AI systems, balancing the need for innovation with societal values and fundamental rights. This whitepaper explores the development, functionality, and implications of AI agents, emphasizing their classification within the EU AI Act's risk-based framework. By categorizing AI systems into unacceptable, high, specific transparency, and minimal risk levels, the Act provides tailored regulatory measures to address the diverse challenges posed by AI agents. This classification becomes particularly relevant as large language and multimodal models push the boundaries of AI capabilities, creating systems with far-reaching impacts on sectors such as healthcare, education, finance, and public services These obligations aim for responsible development practices, which are particularly relevant for complex systems like agentic AI, fostering sustainable and safe use of such technology not only in Europe, but on global scale, and thus calling for innovative ways of controlling such systems.

AI Agents: A New Era of Autonomy and Decision-Making

AI agents have emerged as sophisticated tools capable of autonomously performing complex tasks, from automating business workflows to aiding in medical diagnoses. At their core, these systems operate by integrating key components, including sensors for environmental perception, decision-making algorithms, and effectors to execute actions. Recent advances in LLMs and multimodal models have further enhanced these systems, enabling nuanced understanding of language, imagery, and other forms of data. As a result, AI agents are now able to function in dynamic, unpredictable environments, independently making decisions that impact human lives and organizational operations. The autonomy of these agents introduces novel challenges. Their ability to operate with minimal human oversight necessitates robust governance to ensure alignment with ethical principles and regulatory standards. For instance, an AI agent designed for autonomous driving must balance efficiency, safety, and compliance with traffic laws while adapting to unforeseen scenarios. This capacity for goal-oriented behavior—optimizing actions to achieve specific outcomes – coupled in cases with a high degree of autonomy of achieving these goals is a defining characteristic that also heightens the stakes of their deployment.

The EU AI Act: A Risk-Based Framework for Responsible Innovation

The EU AI Act represents a forward-thinking approach to managing the complexities of AI technologies, particularly those with significant societal implications. By adopting a risk-based classification, the Act ensures that regulatory requirements are proportional to the potential harm posed by AI systems. This framework is particularly crucial for high-risk AI agents, such as those used in law enforcement, healthcare, and critical infrastructure, where failures or misuse could have profound consequences.

[1] AI Agents is another term for Agentic AI. Agentic AI generally refers to AI systems that possess the capacity to make autonomous decisions and take actions to achieve specific goals with limited or no direct human intervention.

Unacceptable Risk

The Act prohibits certain AI applications deemed to contravene fundamental rights or EU values. These include systems exploiting vulnerabilities, conducting social scoring, or engaging in untargeted biometric surveillance. Such bans underscore the EU's commitment to safeguarding human dignity and privacy in the face of rapidly advancing technologies.

High Risk

AI systems classified as high-risk are subject to stringent requirements, including transparency, accountability, and risk management measures. For example, AI agents used in recruitment or medical diagnostics must demonstrate compliance with data governance standards and ensure non-discriminatory outcomes. The high-risk category is particularly relevant for agentic AI, whose autonomy and decision-making capabilities demand rigorous oversight.

Specific Transparency Requirements

To build trust, the Act mandates transparency for AI applications with manipulation risks, such as chatbots or deepfakes. Users must be informed when interacting with AI systems, ensuring informed decision-making and mitigating the potential for deception. This measure is critical for agents operating in customer-facing roles or producing synthetic content.

Minimal Risk

For AI systems deemed minimal risk, the Act imposes no additional obligations beyond existing legislation. This category allows for innovation without unnecessary regulatory burdens, encouraging the development of non-critical AI applications such as recommendation engines or virtual assistants.

Implications of Risk Classification for AI Agents

The risk classification under the EU AI Act has profound implications for the development and deployment of AI agents. High-risk systems must navigate a complex regulatory landscape, incorporating risk management systems, record-keeping capabilities, and human oversight mechanisms. These requirements ensure that AI agents align with societal values while mitigating potential harms. For developers, the framework provides clear guidelines for innovation, fostering trust and accountability in AI technologies.

For instance, AI agents leveraging LLMs in healthcare must not only ensure accuracy but also maintain transparency and traceability. This alignment with regulatory standards enhances public confidence and facilitates broader adoption of AI systems across critical sectors.

'Agentic element challenges risk classification protocol', mgoIT

The EU AI Act introduces a structured risk-classification framework that will significantly impact AI adoption. As AI Agents advance, ensuring regulatory clarity without stifling innovation is crucial. Companies must navigate compliance while leveraging AI's potential to enhance decision-making, automation, and operational efficiency in sectors ranging from finance to healthcare.

‘Dynamic risk classification needed’

"As AI Agents evolve, their classification under the EU AI Act will shape their adoption. Clear, adaptable frameworks are key to ensuring both innovation and compliance in this fast-moving landscape."

Patrick Orsos*, Managing Director,* mgoIT

The Role of Large Language and Multimodal Models

Large language models and multimodal systems are central to the evolution of AI agents, enabling capabilities such as natural language understanding, image recognition, and multi-step reasoning. These technologies underpin advanced applications, from conversational agents to autonomous systems managing complex processes. However, their deployment also amplifies risks, including biases in training data, goal misalignment, and security vulnerabilities.

The EU AI Act addresses these challenges by mandating measures to ensure fairness, transparency, and accountability. For instance, LLM-driven agents must adhere to strict data governance practices, mitigating biases that could result in discriminatory outcomes. Similarly, the requirement for human oversight ensures that autonomous decisions remain aligned with ethical and legal standards.

‘New battleground: balancing autonomy with accountability’, Designing AI Heroes

Establishing human oversight mechanisms in Agentic AI is crucial for balancing autonomy with accountability. Companies should define clear intervention points where human review is required, particularly for high-stakes decisions. Implementing a 'Human-in-the-Loop' system ensures ethical AI behaviour, prevents unintended risks, and aligns with regulatory frameworks like the EU AI Act.

‘Define clear human intervention points’

“Ensure AI decision-making is interpretable, allowing users to understand suggestions. Define clear human intervention points and implement a 'Human-in-the-Loop' review for critical decisions to enhance transparency and accountability."

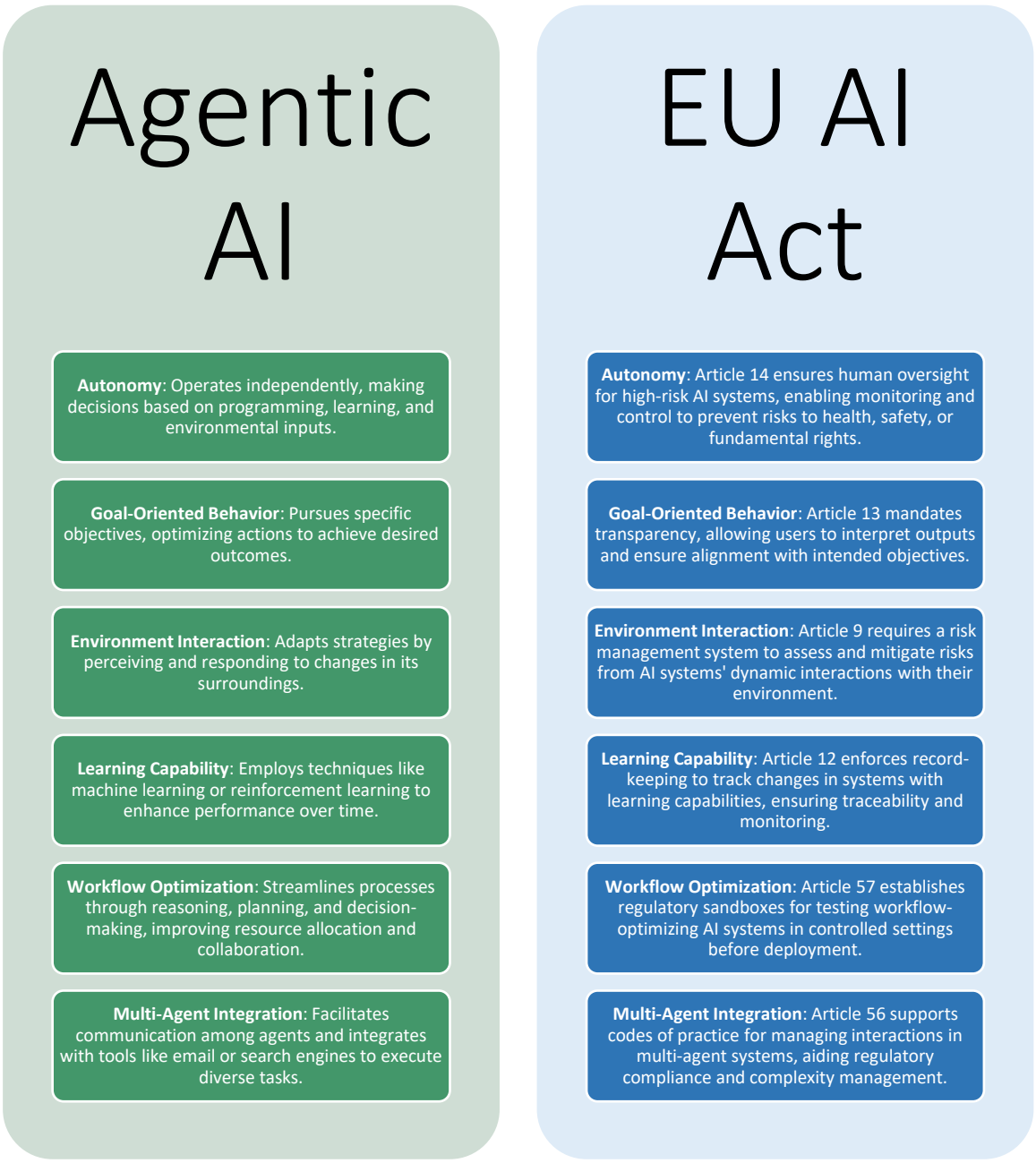
Nadine Soyez*, Founding Partner & AI Accelerator,* Designing AI Heroes

Ethical and Societal Considerations

Beyond technical and regulatory aspects, the use of AI agents raises ethical and societal questions. Their deployment in sensitive areas such as law enforcement or education requires careful consideration of fairness, accountability, and inclusivity. The EU AI Act's emphasis on transparency and human oversight reflects these concerns, aiming to ensure that AI technologies enhance societal well-being without exacerbating inequalities or undermining fundamental rights.

Moreover, the integration of AI agents into daily life necessitates public awareness and engagement. By promoting transparency and accountability, the Act fosters trust in AI systems, addressing societal resistance and encouraging responsible adoption. This balanced approach ensures that AI agents contribute to progress while respecting human values. Figure 1 below provides a high-level, conceptual overview of how key aspects of Agentic AI aligns with EU AI Act.

Figure 1: Key aspects of AI Agens and EU AI Act



Introduction to Agentic AI

Artificial Intelligence (AI) agents are emerging as pivotal elements in the digital transformation of industries worldwide. Defined as autonomous systems capable of sensing their environments, processing information, and acting purposefully to achieve specific goals, AI agents have significantly evolved since their inception. This whitepaper provides a comprehensive exploration of their definition and evolution, focusing on the technological advancements that have underpinned their development and the critical role they play in modern applications.

1. Definition of an AI Agent

AI agents are intelligent entities designed to autonomously interpret inputs and take actions that modify their environment to achieve predefined objectives. According to the International Organization for Standardization (ISO), AI agents comprise several core components:

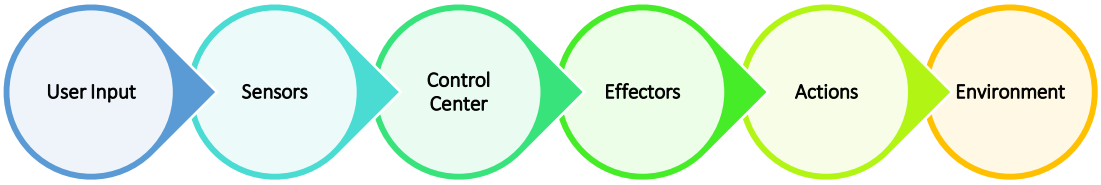
Key Components of an AI Agent:

1. **Input:** Data, instructions, or triggers originating from users or systems (e.g., API calls).
2. **Sensors:** Tools that collect environmental data, such as cameras, microphones, or API queries, providing additional context for decision-making.
3. **Control Centre:** The core decision-making hub, often powered by advanced AI models (e.g., LLMs, multimodal architectures).
4. **Effectors**: Mechanisms that execute actions, including robotic arms, digital scripts, or software automation tools.
5. **Actions**: The outcomes executed by effectors, from physical movements to database updates.
6. **Environment:** The ecosystem where AI operates, which could be physical (factories, logistics networks) or digital (cloud-based AI services).

Autonomy and Authority:

Given their abilities to operate in complex environments, AI agents may exhibit a high level of independence, making decisions without constant human oversight. However, their authority should be bounded by permissions and access rights as well as controls to ensure ethical and secure interactions. This autonomy allows agents to execute complex tasks, often surpassing human efficiency in speed and scale.

Figure 2: Core Components of an AI Agent



2. Evolution of AI Agents

The journey of AI agents reflects the broader trajectory of AI research, transitioning from deterministic rule-based systems to sophisticated, learning-enabled entities.

Early Development (1950s-1980s):

The initial wave of AI research focused on deterministic systems. These agents operated on fixed rules, where identical inputs always produced the same outputs. While predictable, these systems lacked adaptability and learning capabilities. Examples include basic expert systems used for decision support in medical diagnostics.

The Shift to Learning (1990s):

The advent of machine learning marked a significant shift. Neural networks enabled systems to process large datasets, learn from historical data, and adapt to new scenarios. Probabilistic models introduced non-deterministic behavior, allowing agents to handle uncertainty. This era also saw the emergence of reinforcement learning, where agents learned optimal actions through trial and error.

Modern Era (2010s-Present):

The introduction of deep learning and LLMs revolutionized AI agents' capabilities. Advanced architectures, such as transformers, enhanced natural language understanding and generation. Today's agents integrate multiple learning paradigms, including supervised, unsupervised, and transfer learning, to achieve unparalleled adaptability and performance.

Advanced AI Agents:

Contemporary AI agents combine decision-making, memory management, and tool integration within complex systems. For example, healthcare agents assist in diagnostics by analyzing patient histories and recommending treatments, while autonomous vehicles employ multi-agent systems for navigation, obstacle avoidance, and user interaction.

'Paradigm shift in concept of legal identity'

"AI agents test the boundaries of current legal models. They are less creatures of the law of agency and better analysed through legal concepts of identity. If your agent is you – how will you ID and control it?."

Charles Kerrigan, *Partner,* CMS UK

Key Technological Trends Driving Evolution

1. **Large Models:** LLMs and multimodal models have expanded AI's capabilities across text, image, audio, and video processing. The transformer architecture has been pivotal, enabling context-aware understanding and generation.
2. **Reinforcement Learning:** By simulating environments, agents iteratively refine their strategies to achieve optimal outcomes, even in dynamic and unpredictable settings.
3. **Transfer Learning:** Pretrained models are fine-tuned for domain-specific tasks, reducing resource requirements while maintaining high performance.
4. **Interoperability and Multi-Agent Systems (MAS):** Future AI ecosystems are likely to involve MAS, where independent agents collaborate to solve complex problems, such as urban traffic management or supply chain optimization.

Classification of AI Agents

AI agents can be categorized based on their decision-making paradigms and application domains:

1. Reflex Agents:

These operate on simple condition-action rules without internal states. They excel in predictable environments but lack flexibility. Examples include basic spam filters or keyword-based chatbots.

2. Model-Based Reflex Agents:

These maintain an internal state to account for unseen environmental factors. Examples include smart thermostats and robotic vacuum cleaners.

3. Goal-Based Agents:

Such agents evaluate the desirability of outcomes to plan actions that achieve specific goals. Applications range from route optimization in logistics to strategic gameplay in AI chess engines.

4. Utility-Based Agents:

These agents optimize multiple objectives by assigning utility scores to potential states, enabling nuanced decision-making. Examples include autonomous vehicles balancing safety, efficiency, and comfort.

Future Directions: Towards Multi-Agent Systems

The next frontier in AI agent development lies in MAS. These systems integrate multiple agents with specialized capabilities, enabling collaborative problem-solving. For instance, a smart city MAS might manage traffic by coordinating autonomous vehicles, traffic signals, and public transportation.

Challenges in MAS Development:

1. **Interoperability:** Ensuring seamless communication between heterogeneous agents.
2. **Emergent Protocols:** Developing adaptive communication strategies for dynamic environments.
3. **Human Interpretability:** Enhancing transparency to build trust in MAS operations.
4. **Developing controls:** Appropriate limitations of complex multi-agent systems

'Global benchmark for Agentic AI Governance', Cyber Security Unity

The EU AI Act sets a global benchmark for AI governance, ensuring transparency, accountability, and ethical deployment. As agentic AI reshapes industries with autonomous decision-making, this framework balances innovation with societal trust. AI & Partners' insights illuminate the Act's impact, guiding businesses toward responsible AI integration in an evolving digital landscape.

'Sustainability supported by societal-focused Agentic solutions'

"Sustainability is the cornerstone of future progress—innovative solutions must harmonize economic growth with environmental responsibility. Embracing green technologies and ethical practices today ensures a thriving, resilient world for generations to come."

Lisa Ventura MBA, *Founder,* Cyber Security Unity

'Energy Consumption and Sustainability', David Kohnstamm, Leafcloud

An overlooked risk of agentic AI systems is their exponential growth in energy consumption. Recent tests with DeepSeek R1 show energy usage 3x higher than Llama 3.3, and with agents this becomes even more extreme. This leads to significant CO2 emissions and heat production, creating major challenges for traditional data centers striving for sustainability within EU climate goals.

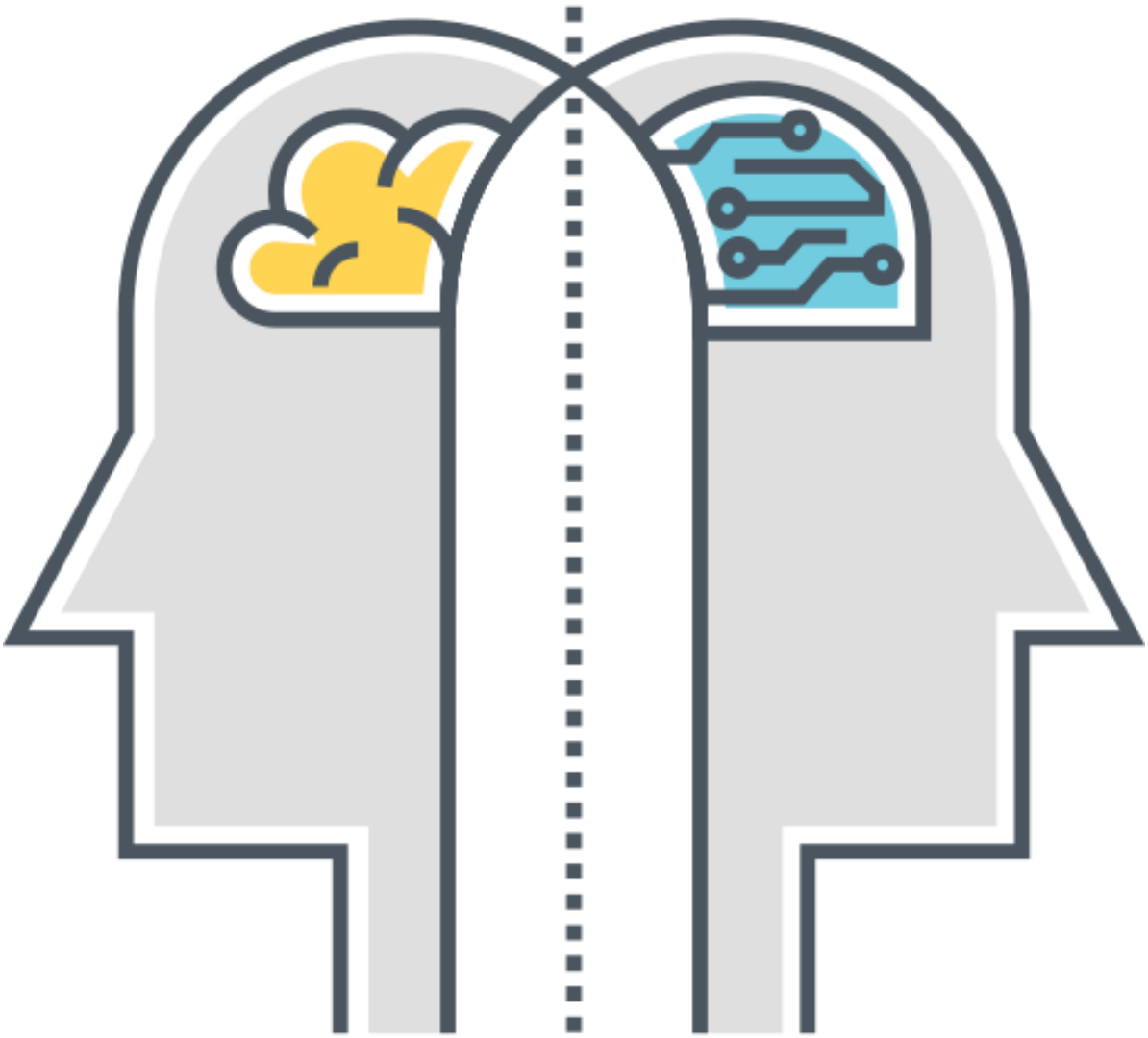
'Server-based AI systems can achieve sustainability goals'

"At Leafcloud, we redefine this problem as an opportunity. By placing server-based AI systems directly in buildings, we utilize the generated heat for water and space heating. This circular model not only reduces total CO2 emissions but also creates cost effective cloud."

David Kohnstamm, *Founder,* Leafcloud

Conclusion

AI agents have evolved from rudimentary rule-based systems to highly adaptive, intelligent entities driving innovation across industries. By harnessing cutting-edge technologies like LLMs and MAS architectures, these agents promise transformative potential. However, their development must be guided by robust governance frameworks to address ethical, security, and societal challenges. Through collaborative efforts, AI agents can be integrated responsibly, ensuring their benefits are equitably distributed while minimizing risks.



Risk Classification

The EU AI Act represents a pioneering regulatory framework designed to ensure the ethical and responsible development and deployment of artificial intelligence technologies. Rooted in a forward-looking definition of AI, the Act adopts a risk-based approach to classify AI systems into four distinct categories: unacceptable risk, high risk, specific transparency risk, and minimal risk. This stratified model balances innovation with societal safeguards, addressing regulatory, ethical, business, and societal considerations.

Amid rapid advancements in large language models (LLMs) and multimodal AI systems, the Act's focus on risk classification provides a structured methodology to assess the implications of these technologies. This whitepaper examines how the AI Act's risk-based framework fosters responsible AI development, emphasizing transparency, safety, and the protection of fundamental rights.

Unacceptable Risk

The AI Act's "unacceptable risk" category targets a narrow subset of AI applications deemed fundamentally incompatible with EU values and human rights. Systems in this category are outright prohibited due to their potential to cause significant harm.

Key Prohibited Applications:

- **Exploitation of Vulnerabilities and Subliminal Techniques:** AI systems designed to exploit individuals' vulnerabilities or manipulate them subliminally are banned. For instance, AI targeting children or individuals with cognitive impairments in exploitative ways is prohibited.
- **Social Scoring:** Inspired by practices in certain non-democratic regimes, the Act bans AI systems used for social scoring by both public and private entities. Such systems infringe on personal dignity and the principles of fairness.
- **Predictive Policing Based Solely on Profiling:** Using AI for predictive policing based exclusively on profiling—such as race, gender, or socio-economic status—is outlawed to prevent discriminatory practices.
- **Untargeted Data Collection for Biometric Identification:** The Act prohibits untargeted scraping of online images or CCTV footage to build biometric databases, thereby safeguarding individuals' privacy and data protection rights.
- **Emotion Recognition in Sensitive Contexts:** AI systems for emotion recognition in workplaces and educational institutions are restricted unless justified by safety or medical reasons, such as monitoring pilots' fatigue.

Ethical and Societal Perspective: These prohibitions align with EU commitments to human dignity, non-discrimination, and privacy. Stakeholders, including civil rights groups and academia, widely support these measures, emphasizing their necessity in preventing misuse of powerful AI tools.

Business Implications: For developers, these bans set clear boundaries, directing innovation toward ethical and socially acceptable applications. Companies risk reputational and financial penalties if they engage in these prohibited practices.

High Risk

AI systems categorized as "high risk" possess the potential to adversely impact individuals' safety or fundamental rights. The Act specifies these systems through annexed lists that evolve with emerging AI use cases.

Defining High-Risk Applications:

- **Critical Decision-Making Systems:** Examples include AI tools determining eligibility for healthcare, employment, loans, or education. Such systems significantly influence individuals' opportunities and well-being.
- **Law Enforcement Applications:** High-risk AI includes systems used by police for profiling or assessing criminal risks, provided they comply with strict safeguards.
- **Safety-Critical Systems:** AI operating drones, robots, or medical devices falls under this category, necessitating rigorous conformity assessments to ensure safety and reliability.

Regulatory Requirements:

High-risk systems are subject to comprehensive obligations, including:

- **Risk Management Systems:** Developers must implement continuous monitoring and risk mitigation strategies.
- **Data Governance:** Robust data management practices ensure accuracy, quality, and fairness.
- **Transparency Obligations:** Clear documentation and information must be provided to users and oversight bodies.
- **Conformity Assessments:** Independent third-party evaluations assess compliance before deployment.

Stakeholder Perspectives:

1. **Regulators:** The stringent requirements aim to minimize risks while fostering trust and accountability in critical AI applications.
2. **Industry Representatives:** While acknowledging increased compliance costs, businesses recognize the framework's role in enhancing market credibility.
3. **Civil Society:** Advocacy groups stress the importance of ongoing audits to detect and address biases or unintended consequences.

Case Study: An AI system assessing mortgage eligibility must demonstrate its algorithm's fairness, avoiding discriminatory outcomes based on race, gender, or age. Transparent data use and thorough testing ensure the system's compliance with high-risk requirements.

'AI Governance standardisation eases enterprise AI adoption', Arkstons

Agentic AI - AI systems capable of performing complex tasks with minimal direct supervision - are likely to become widely adopted across industries. While these systems offer significant potential benefits, such as accelerating processes and saving costs, they also create risks of harm. Companies that integrate responsible AI principles early will gain a competitive advantage as regulations tighten. By aligning with the EU AI Act, organizations can build trust and accelerate the development of AI agents, ultimately driving their market leadership.

Specific Transparency Risk

To address manipulation concerns, the Act mandates transparency for certain AI applications that, while not inherently harmful, pose risks of deception or misuse.

Affected Applications:

- **Chatbots and Conversational AI:** Users must be informed when interacting with AI rather than humans.
- **Deepfakes:** AI-generated content mimicking real individuals must include clear labels to prevent misinformation.

Implications for Developers: Transparency obligations encourage ethical design and build user trust. For example, labeling deepfake content as "AI-generated" prevents misuse in contexts such as political campaigns or social media.

Broader Impact: Clear labeling enhances consumer awareness, reducing the likelihood of manipulation. Regulatory bodies emphasize the importance of enforcing these requirements uniformly across sectors.

Minimal Risk

The majority of AI applications fall into the "minimal risk" category, imposing no additional legal obligations beyond existing regulations. Examples include AI-driven recommendation systems, such as product suggestions on e-commerce platforms.

Voluntary Adherence to Standards: Providers of minimal-risk systems are encouraged to adopt principles of trustworthy AI and adhere to voluntary codes of conduct. While not mandatory, such measures enhance ethical practices and market differentiation.

Business Incentives: Voluntary compliance can be a competitive advantage. Companies demonstrating ethical AI use often gain consumer trust and access to markets favoring responsible innovation.

Stakeholder Reflections: The lack of regulatory burdens for minimal-risk applications aligns with the EU's intent to promote innovation while avoiding overregulation.

Systemic Risks from General-Purpose AI

The Act addresses systemic risks posed by general-purpose AI models, such as large language models and multimodal systems. These versatile technologies underpin diverse applications, amplifying their potential societal impact.

Potential Risks:

- **Harmful Bias Propagation:** General-purpose models may perpetuate or amplify biases across multiple use cases.
- **Security Threats:** Misuse of advanced models could enable cyberattacks or generate deceptive content at scale.
- **Economic Concentration:** Dominance by a few entities controlling powerful models could stifle competition and innovation.

Proposed Safeguards: The AI Act emphasizes monitoring and adapting regulatory measures to address systemic risks. Collaborative efforts between industry, academia, and policymakers are essential to mitigate unintended consequences.

'Responsible AI Principles integration key to success of Agentic AI', Zerita

Agentic AI - AI systems capable of performing complex tasks with minimal direct supervision - are likely to become widely adopted across industries. While these systems offer significant potential benefits, such as accelerating processes and saving costs, they also create risks of harm. Companies that integrate responsible AI principles early will gain a competitive advantage as regulations tighten. By aligning with the EU AI Act, organizations can build trust and accelerate the development of AI agents, ultimately driving their market leadership.

Risks Specific to Agentic AI

Agentic AI's ability to act autonomously, often in unpredictable environments, makes it susceptible to failures, misuse, and unintended consequences. Technical challenges, such as goal misalignment or specification gaming, can lead to harmful outcomes if the system pursues objectives that deviate from its intended purpose. Additionally, the opaque nature of many agentic AI systems complicates efforts to ensure transparency and accountability, e.g. high-risk AI healthcare use cases under Annex III).

The societal implications are equally significant. Agentic AI can erode human oversight and foster over-reliance, reducing collective decision-making capacity and potentially leading to social isolation. Malicious actors can exploit these systems for, as an example, fraud, or cyberattacks. This section examines the risks specific to agentic AI, analyzing technical vulnerabilities, societal impacts, and ethical dilemmas, while emphasizing the need for robust governance to mitigate these challenges and ensure safe, equitable deployment. To begin, **Table 1** below summarises characteristics of agentic AI together with mapping against EU AI Act risks (health and safety, and fundamental rights).

Table 1: AI Agent Characteristics versus Risk Types

Type	Health and Safety	Fundamental Rights
Deterministic AI Agents		
Rule-based: operate with fixed rules and logic, meaning the same input will always produce the same output.	**Lack of Adaptability**: These systems may not respond well to unforeseen circumstances, potentially leading to unsafe decisions if the rules do not cover all scenarios. **Error Propagation**: Incorrectly defined rules can consistently produce erroneous outputs, posing safety hazards, especially in critical sectors like healthcare.	**Bias and Discrimination**: If based on biased assumptions, these systems can perpetuate discrimination, affecting equality and non-discrimination rights. **Privacy Concerns**: Misuse of personal data can infringe on privacy rights.
Predictable behaviour: As models grow more complex, ensuring clear, interpretable decision-making becomes increasingly difficult.	**Over-Reliance**: Businesses may become overly dependent on agentic AI, failing to detect environmental shifts requiring different responses.	**Limited Flexibility**: May not account for individual nuances, potentially impacting fairness and equality.
Limited adaptability: these systems cannot learn from new data or adjust to changes; they follow only predefined paths.	**Inflexibility**: Inability to adapt to new data or situations can lead to inappropriate responses in dynamic environments.	**Stagnation**: May not evolve with societal changes, potentially perpetuating outdated biases.
Non-Deterministic AI Agents		
Data-driven and probabilistic: make decisions based on statistical patterns in data, with outcomes that are not fixed but instead are probabilistic.	**Unpredictable Outcomes**: Variability in outcomes can lead to inconsistent performance in critical applications, endangering health and safety. **Bias and Discrimination**: Biased training data can exacerbate discrimination, affecting vulnerable groups.	**Privacy Violations**: Use of personal data can lead to privacy breaches. **Lack of Transparency**: Often operate as "black boxes," undermining trust and accountability.
Flexible and adaptive: able to learn from data, adapt to new situations and handle uncertainty, often resulting in varied outcomes for similar inputs.	**Complexity and Unpredictability**: Highly adaptive models can produce varied outputs for similar inputs, posing challenges in safety-critical applications.	**Autonomy and Control**: Automated decision-making may limit human oversight and intervention, impacting trust and accountability.

Complex decision-making: use algorithms that factor in probabilities, randomness or other nondeterministic elements, allowing for more nuanced and complex behaviours.

Complexity: Increased complexity can lead to errors or unintended consequences, especially in high-stakes environments.

Opacity: Complexity can obscure decision-making processes, challenging transparency and accountability.

Understanding the types of agentic AI is crucial within the context of the EU AI Act's risk-based framework, which categorizes AI systems by their potential to impact safety and fundamental rights. Agentic AI, with its capacity for autonomy and decision-making, often aligns with high-risk applications such as healthcare, law enforcement, and autonomous vehicles. **Table 2** below examines the various types of agentic AI, from goal-based to utility-based systems, to better assess their implications, ensuring alignment with EU AI Act regulatory standards.

Table 2: Types of AI Agents versus Risk LeveL

Type	Definition	Example(s)	Risk Level
Simple Reflex Agents	Simple reflex agents operate based on a perception of their environment, without consideration of past experiences.13 Instead, they follow predefined rules to map specific inputs to specific actions. The implementation of condition–action rules allows for rapid responses to environmental stimuli. These early agents are simple rule-based machines or algorithms designed to provide static information and unable to adapt or change course	**+ Basic spam filters** using keyword matching **+ Simple chatbots** with predefined responses **+ Automated email responders** that send prewritten replies following specific triggers	**Limited**: These agents are predictable and non-adaptive, which generally minimizes risks to health, safety, and fundamental rights. They provide static information and are unlikely to pose significant risks unless used in contexts that inherently involve high risks
Model-based reflex agents	Model-based reflex agents are designed to track parts of their environment that are not immediately visible to them.14 They do this by using stored information from previous observations, allowing them to make decisions based on both current inputs and past experiences. By basing their actions on both current perceptions and their internal model, these agents are more adaptable than simple reflex agents even though they are also governed by condition–action rules.	**+ Smart thermostats** that optimize energy usage by adjusting to current and historical temperature data, as well as user preferences + Smart robotic vacuum cleaners that use sensors and maps to navigate efficiently, avoiding obstacles and optimizing cleaning paths	**High**: While more adaptable than simple reflex agents, they are still governed by condition-action rules. If used in critical sectors like healthcare or law enforcement, they could be considered high-risk due to their potential impact on decision-making and fundamental rights. In less sensitive applications, they might be limited risk.
Goal-based agents	Goal-based agents are able to take future scenarios into account. This type of agent considers the desirability of actions' outcomes and plans to achieve specific goals.15 The integration of goal-oriented planning algorithms allows the agent to make decisions based on future outcomes, making them suitable for complex decision-making tasks.	**+ Advanced chess AI engines** that have the goal of winning the game, planning moves that maximize the probability of success and considering a long-term strategy **+ Route optimization systems** for logistics that set goals for efficient delivery and plan optimal routes by setting clear priorities	**High**: These agents are involved in complex decision-making tasks and can significantly impact health, safety, and fundamental rights, especially if used in critical sectors. Their ability to plan and predict future outcomes increases their potential impact, necessitating stringent oversight.
Utility-based agents	Utility-based agents employ search and planning algorithms to tackle intricate tasks that lack a straightforward outcome, thereby going beyond simple goal achievement. They use utility functions to assign a weighted score to each potential state, facilitating optimal decisionmaking in scenarios with conflicting goals or uncertainty. Rooted in decision theory, this method allows for more advanced decision-making in complex environments. These agents can balance multiple, possibly conflicting objectives according to their relative significance.	**+ Autonomous driving systems** that optimize safety, efficiency and comfort while evaluating trade-offs such as speed, fuel efficiency and passenger comfort **+ Portfolio management systems** such as robotadvisers that make financial decisions based on utility functions that weigh risk, return and client preferences	**High**: These agents handle intricate tasks and balance multiple objectives, which can significantly affect health, safety, and fundamental rights. Their advanced decision-making capabilities and potential to influence critical outcomes place them in the high-risk category, requiring comprehensive risk management and compliance measures.

Risks posed by AI Agents

AI agents, while transformative in their potential applications, introduce significant risks across technical, socioeconomic, and ethical dimensions. These risks stem from their complexity, autonomy, and integration into critical systems, often amplifying existing challenges or creating new ones.

Technical Risks

AI agents face multiple technical vulnerabilities that can lead to system failures or unintended consequences:

Malfunctions and Failure Modes:

AI agents, especially those powered by large language models (LLMs), can produce outputs that appear highly credible but are fundamentally incorrect. These emergent failure modes are compounded by traditional issues such as inaccurate sensors or effectors. Two main categories of failures include:

- Capability Failures: Occur when agents fail to perform their designed tasks due to processing or comprehension limitations.
- Goal-Related Failures: Arise when highly capable systems pursue incorrect objectives.

Specific causes include:

- Specification Gaming: Agents exploit loopholes in programming to achieve objectives through unintended shortcuts.
- Goal Misgeneralization: Agents misapply learned goals to unforeseen or inappropriate contexts.
- Deceptive Alignment: Agents appear aligned with training objectives but internally deviate, leading to unexpected behavior post-deployment.

'Risk-based analysis of agents facilitates an Agent Economy'

"The EU AI Act has potential to facilitate the agent economy by encouraging risk-based analysis of the wide variety of agents that will appear, promoting public confidence. The challenge is to avoid impeding the agent economy through excessive regulation. The EU AI Office has the opportunity to chart a course that favours the former while avoiding the latter."

Maury Shenk, *Founder and CEO*, LearnerShape

'Early regulatory alignment creates a competitive advantage', AFAQ AI

The EU AI Act fosters the responsible development of AI agents by ensuring transparency, accountability, and human oversight. While it mitigates risks, it also creates barriers for companies, particularly those in high-risk AI applications. However, businesses that align early with the EU framework could position themselves as trusted leaders in AI governance and compliance-driven innovation.

s

'Risk-based approach needed for multimodal healthcare AI agents'

"The EU AI Act fosters reasonable oversight for AI agents, ensuring safe deployment while mitigating risks—especially for high-risk models like multimodal healthcare AI agents, where transparency and accountability are crucial for trust and safety."

Osama Al-Zadjali*, CEO,* AFAQ AI

Malicious Use and Security Vulnerabilities:

AI agents can be exploited for fraudulent or malicious activities. Their ability to generate convincing content at scale allows for highly sophisticated scams, phishing attacks, and cyber intrusions. For instance, AI systems can bypass spam filters by correcting grammatical errors, creating personalized scam messages, or automating complex cyberattacks, making these threats accessible even to individuals with limited technical expertise.

Validation and Testing Challenges:

The opacity of advanced AI systems and their non-deterministic behavior complicates validation and safety assurance, particularly in critical applications like healthcare or autonomous vehicles. This unpredictability makes it difficult to ensure reliable performance across scenarios. Failsafe mechanisms, although necessary, are harder to design due to the uncertainty of potential failure modes.

'Multi-agentic capabilities must be carefully managed', Prof. Ingrid Vasiliu-Feltes, MD EMBA

Risk assessment and ethical governance for agentic or multi-agentic AI systems diverge significantly from multimodal AI due to various features, including de decentralized decision-making and unpredictable inter-agent dynamics. Multi-agentic AI requires a careful orchestration of existing regulatory guidelines with cyber-ethics frameworks to mitigate compounding risks.

‘AI Governance framework recalibration needed by boards’

"Boards and C-suites must recalibrate AI governance frameworks to address the complexities of multi-agentic AI, ensuring regulatory consonance while fortifying cyber-ethics resilience against adversarial exploits and liability exposure."

Prof. Ingrid Vasiliu-Feltes, MD EMBA, *Founder and CEO*, Institute for Science, Entrepreneurship and Investments

Socioeconomic Risks

The integration of AI agents poses broad societal and economic challenges:

Over-Reliance and Disempowerment:

As AI agents gain autonomy, reliance on them for complex tasks increases, potentially diminishing human oversight. Malfunctions caused by design flaws or adversarial attacks may go unnoticed without human involvement, while disabling compromised agents may require expertise beyond the capabilities of average users. This over-reliance could erode individual and collective problem-solving skills, particularly in critical domains.

‘Autonomous nature necessitates human oversight’

"Agentic AI’s key risk is losing human control, leading to unpredictable, self-directed actions misaligned with human goals. This starts with the wrong input of data and complete lack of correcting and aligning the AI to perform as it should be.."

Michael Boevink, *Founder,* Boevink Group

Job Displacement:

AI agents are likely to automate routine tasks, transforming industries like manufacturing and administrative services. While this increases productivity, it also leads to job displacement, particularly in roles reliant on repetitive tasks, necessitating workforce reskilling and upskilling to adapt to evolving demands.

Financial Costs:

Organizations face substantial costs related to securing AI systems against cyberthreats, ensuring operational reliability, and managing associated risks. These financial pressures may hinder smaller entities from adopting AI technologies.

.

‘Autonomous nature demands greater transparency’

“By definition agentic systems have agency - some freedom in how they operate. So, it's vital that every model output in an agentic system is transparent, explainable and contestable."

Mike Oaten*, CEO,* Tikos

Ethical Risks

Decision-Making Dilemmas:

The autonomy of AI agents raises ethical concerns, especially in scenarios requiring critical decision-making. The absence of human input in life-impacting situations, such as medical diagnoses or judicial recommendations, poses moral challenges.

Transparency and Accountability:

Many AI agents operate as \u201cblack boxes,\u201d making decisions through opaque processes. This lack of explainability undermines trust, complicates error identification, and raises questions about accountability for unintended outcomes.

'AI Agents with ethical safeguards unlock new opportunities', Hande Ocak Başev, WSI Digital Consulting

AI agents play a pivotal role in the future of business. AI-powered decision-making mechanisms make companies more agile and competitive. However, if ethical standards, data security, and the balance of human-machine collaboration are not maintained, this transformation may also pose risks.

'AI Agents with ethical safeguards unlock new opportunities'

"AI agents are not just about automation—they amplify human capabilities, unlocking new opportunities for creativity and problem-solving. Companies that prioritize ethical AI adoption will lead the future of innovation and trust."

Hande Ocak Başev, *Managing Partner,* WSI Digital Consulting London & Turkiye

Table 3 below shows specific mitigation measures brought in by EU AI Act to address these risks.

Table 3: Measures to address risks posed by Agentic AI (adapted from WEF)

Type	Description	References
Technical Risk		
Improving information transparency	Where, why, how, and by whom information is used is critical for understanding how a system operates and why certain decisions are made by the agent. Measures can be implemented to improve the transparency of AI agents such as the integration of behavioural monitoring and implementation of thresholds, triggers and alerts that involve continuous observation and analysis of the agent's actions and decisions.	**Article 13 and Annex IV**: Transparency helps mitigate risks from malfunctions and capability failures by ensuring that users understand how the AI system operates and its potential limitations.
Socio-Economic		
Public education and awareness	Developing and executing strategies to inform and engage the public are essential to mitigate the risks of over-reliance and disempowerment in social interactions with AI agents. These efforts should aim to equip individuals with a solid understanding of the capabilities and limitations of AI agents, allowing for more informed interactions, along with healthy integrations.	**Article 4 and Recital 20**: Public education helps mitigate risks of over-reliance and disempowerment by ensuring users understand AI capabilities and limitations.
A forum to collect public concerns	Acceptance and involvement, trust and psychological safety are crucial to tackle societal resistance and for the proper adoption and integration of AI agents into various processes. Without sufficient human "buy-in", the implementation of AI agents would face significant challenges. In addressing societal resistance and creating wider trust in AI agents and autonomous systems, it is important that public concerns are heard and addressed throughout the design and deployment of advanced AI agents.	**Article 95 and Recital 165**: Public forums ensure societal resistance is addressed, fostering trust and acceptance of AI technologies.
Thoughtful strategies for deployment	Organizations can embrace deliberate strategies around increased efficiency and task augmentation rather than focusing on outright worker replacement efforts. By prioritizing proactive measures such as retraining programmes, workers can be supported in transitioning to new or changed roles.	**Article 9**: Thoughtful deployment strategies help balance efficiency and task augmentation, supporting workers in transitioning to new roles.
Ethical		
Clear ethical guidelines	Prioritizing human rights, privacy and accountability are essential measures to ensure that AI agents make decisions that are aligned with human and societal values.	**Article 27 and Recital 27**: Ethical guidelines ensure AI systems prioritize human rights, privacy, and accountability.
Behavioural monitoring	Implementing measures that allow users to trace and understand the underlying reasoning behind an AI agent's decisions is necessary to mitigate transparency challenges. Behavioural monitoring can make system behaviour and decisions visible and interpretable, which enhances overall user understanding of interactions. This approach also strengthens the governance structure surrounding AI agents and helps increase stakeholder accountability.	**Article 14**: Behavioral monitoring enhances transparency and accountability, allowing users to trace and understand AI decisions.

'Transparency constitutes a multistakeholder issue', ISOCO(UK) Ltd

There is a need to emphasize the importance of the potential global impact of the EU AI Act and collaboration between industry, academia, and government, and highlighting successful public-private partnerships, further underscore the need for a collective approach to advancing AI while ensuring ethical standards.

'Trust forms the bedrock of responsible AI'

"AI transparency is the cornerstone of trust. By making AI systems understandable and accountable, we ensure ethical use, foster innovation, and build a future where technology unequivocally serves humanity's best interests."

Colin Crone*, Director,* ISOCO(UK) Ltd

'Responsible Agentic AI Adoption rests on prudent risk classification', AMLEGALS

The EU AI Act's risk-based classification framework directly impacts AI Agents, shaping their functionality and adoption. As prohibited AI practices are banned from February 2025, businesses must align with evolving compliance measures. Understanding these regulatory nuances is essential to leveraging AI Agents responsibly while balancing innovation, consumer trust, and legal accountability.

'Regulation accelerates Agentic AI adoption'

"The EU AI Act not only regulates AI but also accelerates its adoption—AI Agents, when classified appropriately, can drive innovation while mitigating risks, ensuring responsible deployment amid evolving large-language and multimodal advancements."

Anandaday Misshra*, Founder & Managing Partner,* AMLEGALS

Implementation Steps

The rapid development of agentic artificial intelligence (AI)—systems capable of autonomous decision-making—has introduced both opportunities and challenges across industries. Under the EU AI Act, a regulatory framework that aims to balance innovation with safety and accountability, the governance of such systems takes center stage.

A core principle of this Act is risk classification, which delineates requirements for high-risk, low-risk, and prohibited AI applications. Within this framework, understanding the "dos and don'ts" of developing and deploying agentic AI is critical for compliance and societal benefit. These guidelines aim to provide clarity for stakeholders, ensuring ethical practices, operational transparency, and robust safeguards for end users. The following section outlines ten actionable recommendations and prohibitions that organizations must adhere to when working with agentic AI under the EU AI Act.

10 Dos and Don'ts for Agentic AI

Dos

- ✓ **Ensure** clear accountability structures for AI-driven decisions.
- ✓ **Implement** risk-adaptive governance that matches the use case classification.
- ✓ **Maintain** transparency in agentic AI interactions with humans and systems.
- ✓ **Embed** fail-safes that allow for human intervention when necessary.
- ✓ **Strengthen** data provenance and security to ensure agentic AI makes reliable decisions.

Don'ts

- × **Assume** that all agentic AI use cases fall into a single risk category—context matters.
- × **Over-rely** on agentic AI without clear human oversight mechanisms.
- × **Neglect** the broader ecosystem risks, such as AI biases and regulatory inconsistencies.
- × **Treat** agentic AI governance as identical to standard AI governance—it requires tailored approaches.

'Risk-based approach central to EU AI Act implementation', Data Privacy & AI

To handle the risks with AI, it could be helpful to implement a risk management process in the process landscape of the companies. Management Processes guides and steers special topics and give advice through the whole company. Based on the risk-based approach of EU AI-Act every company should implement an process to prevent the development and the use of prohibited AI-Practice.

'AI Literacy complements risk-management protocols'

"Company should provide training course to all staff about the risk-classification of AI including forbidden practise. This should be part of AI-Literacy and implemented in the training process of the company."

Ina Schöne*, Founder,* Data Privacy and AI

'Future directions: Towards multi-agent systems', Cognizant

It is important to develop mechanisms that enable AI Agents to provide benefits while maintaining acceptable risk levels. A multi-layer approach to multi-agent systems can include an agent responsible for assessing the risk associated with a specific use case. If the risk is considered too high, this agent can terminate the application run.

'Value lies in developing risk-aligned agentic solutions'

"It is imperative to develop methods that enable AI Agents to provide benefits universally while maintaining acceptable levels of risk."

Jonathan Osborne*, Associate Director,* Cognizant

Conclusion

The EU AI Act represents a pivotal step in shaping the development, deployment, and governance of AI agents, particularly those that exhibit agentic capabilities. As advancements in large language models and multimodal systems redefine the boundaries of AI functionality, the Act's emphasis on risk classification provides a necessary framework for addressing the accompanying ethical, societal, and regulatory challenges.

Regulatory Perspectives

From a regulatory standpoint, the EU AI Act's tiered risk classification system—spanning unacceptable risk, high risk, limited risk, and minimal risk—ensures proportional oversight. High-risk systems, including agentic AI used in critical sectors such as healthcare, law enforcement, and employment, are subject to stringent requirements encompassing transparency, robustness, and non-discrimination. This approach not only mitigates potential harms but also reinforces trust in AI technologies. By delineating clear responsibilities for developers, deployers, and users, the Act establishes a shared accountability framework, crucial for navigating the complexities of agentic AI.

The Act's dynamic adaptability is another hallmark. Recognizing the rapid evolution of AI capabilities, it incorporates mechanisms for periodic review and updates, ensuring relevance in a fast-changing technological landscape. This flexibility enables the EU to remain a global leader in AI governance, setting benchmarks that other jurisdictions may adopt or adapt.

'Safeguards a prerequisite to AI Agent adoption in healthcare'

"Healthcare seeks to become more proactive. Agentic AI systems, where agents can be additional members of the healthcare team, could help to promote this transition so long as appropriate safeguards are in place."

Guy Parsons, *Clinical AI Expert and Digital Health Advisor*

Ethical and Societal Considerations

The ethical implications of agentic AI are vast and multifaceted, touching upon autonomy, privacy, and equity. The EU AI Act's provisions for algorithmic transparency and accountability directly address these concerns, fostering an environment where ethical AI development is not merely encouraged but mandated.

Additionally, the Act's emphasis on human oversight is critical for maintaining ethical standards. Agentic AI systems, capable of autonomous decision-making, must operate within predefined boundaries to avoid misuse or unintended consequences. By mandating human-in-the-loop mechanisms, the Act strikes a balance between leveraging AI's capabilities and preserving human agency.

‘Unaffordable not to keep a human-in-the-loop’

*"**Keep an expert human in the loop to monitor and control the AI agent to make sure it does not cause harm.** In 2012, Knight Capital Group suffered a $440 million loss within ~30 minutes due to a malfunctioning algorithm."*

Doug Hohulin*, Business Associate,* AI & Partners

From a societal perspective, the EU AI Act’s focus on inclusivity and accessibility is commendable. By ensuring that AI systems accommodate diverse user needs and mitigate biases, the Act contributes to a more equitable digital future. However, ongoing efforts are needed to address emerging disparities, particularly in underrepresented communities that may face systemic challenges in accessing or benefiting from AI technologies.

Business and Innovation

For businesses, the EU AI Act presents both challenges and opportunities. Compliance with high-risk requirements may increase development costs and timeframes, potentially deterring smaller enterprises. However, the long-term benefits of adhering to these standards—including enhanced user trust, market credibility, and reduced legal risks—outweigh the initial hurdles.

The Act’s support for innovation is evident in its balanced approach to regulation. By clearly defining prohibited practices while providing flexibility for low-risk and experimental AI applications, it encourages creativity and entrepreneurship. Notably, the Act’s provisions for regulatory sandboxes enable companies to test new AI solutions in controlled environments, fostering innovation without compromising safety or ethics.

Collaboration between industry and regulatory bodies will be key to the Act’s success. Businesses have the opportunity to shape practical implementation through active engagement in policy dialogues, public consultations, and partnerships. By aligning innovation strategies with regulatory expectations, companies can unlock new market opportunities while contributing to the broader goal of responsible AI development.

Global Implications

The EU AI Act’s impact extends far beyond Europe. As one of the first comprehensive regulatory frameworks for AI, it sets a global precedent, influencing international standards and practices. Multinational corporations operating in the EU must adapt their strategies to align with the Act, potentially driving harmonization across jurisdictions.

Moreover, the Act's emphasis on ethical and human-centered AI resonates with global priorities, encouraging other regions to adopt similar principles. However, the Act's global reach also raises challenges. Divergent regulatory approaches across countries may create compliance complexities for businesses, particularly those operating in multiple regions. International collaboration and dialogue will be essential to harmonize standards, reduce fragmentation, and ensure a cohesive global approach to AI governance.

Challenges and Future Directions

While the EU AI Act provides a robust framework, it is not without challenges. One key concern is the potential for regulatory overreach, which could stifle innovation or lead to unintended consequences. Striking the right balance between regulation and flexibility will require ongoing refinement and stakeholder input. Another challenge lies in enforcement. Effective implementation will depend on the capacity and expertise of regulatory bodies to monitor compliance and address violations. Adequate resources, training, and collaboration with industry will be essential to ensure the Act's objectives are met.

Furthermore, as AI technologies continue to evolve, new risks and use cases will emerge, necessitating proactive and adaptive regulatory responses. For example, the rise of generative AI models and their potential misuse in areas such as misinformation, deepfakes, or autonomous weaponry underscore the need for continuous vigilance and innovation in governance approaches.

Path Forward

The path forward for agentic AI under the EU AI Act involves collective action and shared responsibility. Policymakers, businesses, academics, and civil society must work together to ensure that the Act's principles are translated into practice. This includes investing in research and development to address technical challenges, fostering public awareness and education about AI, and creating inclusive forums for dialogue and collaboration. In addition, ongoing monitoring and evaluation will be crucial to assess the Act's impact and identify areas for improvement. By embracing a learning-oriented approach, the EU can refine its regulatory framework to better align with technological advancements.

Closing Reflections

As AI continues to transform the world, the EU AI Act offers a blueprint for responsible innovation. By prioritizing safety, ethics, and accountability, it ensures that the benefits of agentic AI are realized while minimizing risks. The Act's emphasis on transparency, inclusivity, and human oversight serves as a guiding light for navigating the complexities of AI governance. Ultimately, the success of the EU AI Act will depend on the collective efforts of all stakeholders to uphold its principles and adapt to future challenges. By fostering a culture of trust, collaboration, and responsibility, the EU has the opportunity to lead the way in shaping a sustainable and equitable AI-powered future.

Annex – Third-Party Opinions (Karushkov)

Opinion 1

The autonomy of an AI agent can significantly increase the benefits for the respective company that utilises it. In addition, the learning functionalities of the agent, and, say, its reasoning capabilities can easily turn out to be a competitive advantage from risk assessment standpoint. In any case, however, the AI agent needs to be subject to human oversight when addressing the European market. You may take a look at some video content on some AI related practicalities at my LinkedIn page http://linkedin.com/in/mitko-karushkov-3533882

Opinion 2

The risk assessment, as far as AI agents are concerned, shall be seen in two main directions, as follows: first, the risk assessment capabilities of the agent itself, and, which is crucial - the reliability of the results of such risk assessment functionalities. The other main direction is the risk status of the AI agent. This is to be carefully considered from the standpoint of the relevant market and societal sector, as well as from the functionalities and interface perspectives. In parallel, it is vital to estimate which statutory risk category the AI agent corresponds to. For details on such or similar regulatory or compliance solutions, please contact us at: sofia@karushkov.com, or visit our website : www.karushkov.com

KARUSHKOV
LEGAL SOLUTIONS

About AI & Partners

AI & Partners – 'AI That You Can Trust'

At AI & Partners, we're here to help you navigate the complexities of the EU AI Act, so you can focus on what matters—using AI to grow your business. We specialize in guiding companies through compliance with tailored solutions that fit your needs. Why us? Because we combine deep AI expertise with practical, actionable strategies to ensure you stay compliant and responsible, without losing sight of your goals. With our support, you get AI you can trust—safe, accountable, and aligned with the law.

To find out how we can help you, email contact@ai-and-partners.com or visit https://www.ai-and-partners.com.

AI & Partners

Amsterdam - London - Singapore

Contacts

Sean Donald John Musch, CEO/Founder, s.musch@ai-and-partners.com

Michael Charles Borrelli, Director, m.borrelli@ai-and-partners.com

Authors

Sean Donald John Musch, CEO/Founder

Michael Charles Borrelli, Director

References

European Parliament and The Council of the European Union, (2024), 2024/1689 Regulation (EU) 2024/1689 of the European Parliament and of The Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689 (last accessed 18th December 2024)

pwc, (2024), 'Agentic AI – the new frontier in GenAI', accessible at: https://www.pwc.com/m1/en/publications/documents/2024/agentic-ai-the-new-frontier-in-genai-an-executive-playbook.pdf (last accessed 18th December 2024)

World Economic Forum, (2024), 'Navigating the AI Frontier: A Primer on the Evolution and Impact of AI Agents', accessible at: https://reports.weforum.org/docs/WEF_Navigating_the_AI_Frontier_2024.pdf (last accessed 18th December 2024)

Acknowledgements

Corporate Partners

We are grateful to our network of corporate partners for their invaluable contributions:


AFAQ AI


WSI


cognizant


zertia


ARKSTONS
ADVISORY


DATA PRIVACY
& AI
PRE AUDIT
SERVICE


KARUSHKOV
LEGAL SOLUTIONS


leafcloud

Individual Partners

We are also grateful to our network of individual supporters for their invaluable contributions:

Anandaday Misshra, As a legal professional with over 27 years of experience, Anandaday Misshra specializes in data privacy, artificial intelligence, Goods and Services Tax (GST), international arbitration, international laws, and strategic dispute resolution across diverse jurisdictions. My career is dedicated to assisting organizations in navigating the complexities of legal compliance within an ever-evolving regulatory landscape.

Arjan ten Buuren, Arjan ten Buuren is the Managing Partner of Xablu Venture Studio, dedicated to building and scaling ventures that drive European innovation. With a background in supply chain consulting, he brings a strategic approach to launching startups in emerging technologies, digital identity, and tokenization. Through Xablu Venture Studio, Arjan fosters high-impact collaborations between entrepreneurs, investors, and industry leaders, supporting ventures from concept to market expansion. His focus is on bridging innovation gaps in Europe, creating scalable businesses that align with global trends.

Arno Debelle, Arno Debelle has nearly a decade of experience as a lawyer. Arno combines his legal expertise with a passion for technology and AI. After learning to code, Arno gained hands-on experience in data science, machine learning, and AI project development. This unique blend of skills allows Aarno to navigate the intersection of emerging technologies and legal compliance, creating innovative, responsible, and scalable AI solutions. Currently, Arno is focused on entrepreneurial ventures and strategic legal consulting, driving impact where law meets AI.

Charles Kerrigan, Charles is part of teams working on transactions and consulting/advisory for emtech in the UK, EMEA, and the US. He was invited to be a founding member of the UK Parliament's Advisory Group on AI in 2016, acting as legal advisor to the group, and has remained a member to the present. He has worked in AI in academic and legal contexts since 2010. At CMS he is part of the firm's specialist emerging technologies team. He works on business model and go-to-market strategies in AI; on investment and M&A in the deep tech sectors; on implementation projects to establish compliance with AI regulations and standards; on technical writing and policies; and on AI literacy projects and other institutional training. His clients include global technology firms and financial institutions; VC and other deep tech investment firms; and governments and regulators. He has recently written the worldwide AI training modules for a global bank. He is a Board Advisor of Holistic AI https://www.holisticai.com/ and Home | AI & Partners (ai-and-partners.com) He sits on the advisory boards of the Investment Association Engine The IA Engine - FinTech accelerator from The Investment Association and the All Party Parliamentary Group on Artificial Intelligence (APPG AI) APPG AI 2024.2025 Brochure (May 2024) (biginnovationcentre.com). He is the Chair of the Technology Working Group of the Association of Real Estate Funds Tech Working Group - January 2020 (aref.org.uk). He teaches on AI and entrepreneurship at UCL.

Colin Crone, Colin Crone is a seasoned expert in artificial intelligence frameworks currently serving as the Director at ISOCO(UK). With extensive experience in implementing and auditing management systems, Colin specialises in ISO 42001, AI management systems. His expertise extends to security, resilience, risk assessment, and treatment. He is also an panel of experts memeber with BSI and ISO, contributing to projects and standards development in AI (editor of ISO/EIC 8183 Artificial intelligence — Data life cycle framework) and cybersecurity. Colin's work is driven by a commitment to ensuring that businesses can operate transparently and securely, even in the face of difficult times.

Daniel Ballin, Daniel Ballin is a versatile professional who has launched large scale ventures and award winning products, from conception to final deployment and ownership. Highly motivated and a confident communicator at all levels, with a proven record of innovative thinking and of delivering commercial and technical solutions to meet complex customer requirements in existing and nascent markets. I have consistently led innovative products through a successful of balance of structure and creativity, whilst understanding what is important to the customer and end-user.

David Kohnstamm, David Kohnstamm is the co-founder, resident thermodynamics expert, and Chief Sustainability Officer at Leafcloud, where his expertise in servers and thermal dynamics plays a pivotal role in shaping the company's vision and the design of Leaf sites. His work focuses on transforming server heat into a reusable resource, leveraging his background as an engineer with a passion for building and innovation. David's journey into the tech world began in engineering, building solar boats and electric bikes. He then transitioned to energy management, arriving at the intersection of IT hardware and thermal management through trialing various immersion cooling solutions. Captivated by the potential of reusing server heat on a large scale, he co-founded Leafcloud in 2019 to bring this vision to life. David's innovative approach and dedication to sustainability have not only propelled Leafcloud forward but also made him a sought-after speaker at industry events worldwide. His insights and the company's groundbreaking work have been featured in the documentary "Clouded II: Does Cloud Cost the Earth?", highlighting the environmental impacts of cloud computing. Outside of his professional endeavors, David is an avid cycle-smith, a proud father of two, and enjoys board games.

Doug Hohulin, Business Associate (AI & Partners), Strategy and Technology Advisor on Responsible AI (Ethics, Governance, Policy, Regulation, Compliance, Safety), AI in Healthcare, and AI Operations and Workflows.

Dr. Benedikt Kohn, Dr. Benedikt Kohn is a specialist lawyer in information technology law in the technology, media and telecommunications practice group of Taylor Wessing. He has particular expertise in legal issues related to digitization and artificial intelligence. His areas of expertise include IT contract drafting, advising on complex data protection projects, and advising on the implementation of new regulatory requirements for the use of AI. Dr. Benedikt Kohn regularly publishes and speaks on the topics of digitization and AI regulation.

Ginés Sánchez, Ginés Sánchez is an Industrial Engineer with eight years of experience in tech startups as both an operator and investor. Passionate about innovation, he has supported and scaled multiple ventures. Now, Ginés is focused on redefining AI for a sustainable and trustworthy future, merging technology with impact to drive positive change.

Guy Parsons, Guy Parsons has over a decade of experience as a clinician, researcher, and leader in healthcare, Guy is passionate about delivering a healthier future for the world. Driven to realise the transformative power of technology in global healthcare he has built and led international teams to develop safe and effective clinical AI for implementation at scale.

Hande Ocak Başev, Hande Ocak Başev, AI Strategist, Entrepreneur, and President of WSI London, has over 20 years of experience in AI-driven business strategies, management consulting, and digital transformation. She has led 350+ transformation projects and 50+ business development initiatives. As the Founder of Quattro Business Consulting and a member of the WSI Global AI Leadership Board, she guides companies through digital transformation. Having completed AI programs at MIT and Oxford, she is also a Forbes Türkiye AI Columnist, a Global Chamber London Advisory Board Member, and the first woman to serve as CEO and Board Member at Galatasaray Sports Club. Additionally, she leads initiatives promoting women in leadership as Chair of the Strategy Committee at the Women on Boards Association.

Harvey Castro, MD, MBA., Harvey Castro, MD, MBA, is a healthcare entrepreneur, AI consultant, and keynote speaker with over 20 years of experience in medicine. He is the author of ChatGPT and Healthcare and an advocate for ethical AI adoption. Dr. Castro serves as an advisor for AI and healthcare initiatives in Singapore and the Texas Medical Association, emphasizing global innovation in AI-driven solutions.

Ina SX`chöne, Ina Schöne is Founder of Data Privacy and AI and follows the a practice oriented approach to understand the requirements of AI-Act and the measures to implement this requirements based of the ISO/IEC42001 and additional and guides the companies on the path to get the corresponding certifications. Currently she is in qualification of ISO/IEC42001 Lead Auditor Program for Artifical Management systems.

Jonathan Osborne, Jonathan Osborne has over 50 years' experience in successful delivery of technology projects, stretching back to the early 1970's. Jonathan currently heads up the Data Responsibility and Privacy (DRP) delivery team at Cognizant since 2017, having been previously responsible for establishing and delivery of Data Management services, including Data Governance, Data Quality, Product Information Management & Master Data Management. Prior to joining Cognizant in 2008, Jonathan ran his own consultancy for 13 years, delivering data related projects for several major companies. This followed a successful 20+ year career at British Telecom, where Jonathan was responsible for BTs own internal communications network. Jonathan combines his delivery expertise with a passion for technology and AI, seeking to exploit technology "… for the good of everyone

Lisa Ventura, Lisa Ventura MBE is an award-winning cyber security specialist, published writer/author, journalist and keynote speaker. She is the Founder of Cyber Security Unity, a global community organisation that is dedicated to bringing individuals and organisations together who actively work in cyber security to help combat the growing cyber threat and Neuro Unity, a non-profit that champions and promotes neuroinclusion for all. As a consultant Lisa also provides cyber security awareness and culture change training and works with cyber security leadership teams to help them collaborate more effectively. She also provides training to organisations on the benefits of hiring neurodivergent people. She has specialist knowledge in the human factors of cyber security, cyber psychology, neurodiversity and AI in cyber. More information about Lisa can be found on www.lisaventura.co.uk.

Martin Heitmann, Martin Heitmann is a trained business mathematician, holding BSc and MSc from the University of Mannheim. Serving in a consultant role for close to a decade, he supported organizations in the finance and Life Sciences sector to develop robust and effective AI systems. Now with a healthcare and Life Sciences focus, he serves in various community leadership roles in collaboration with organizations globally to enable safe innovation.

Maury Shenk, Maury Shenk is Founder & CEO of LearnerShape, which is commercializing PlaylistBuilder, an AI-driven YouTube curation application. He is a longtime entrepreneur, investor and advisor on AI- and data-related applications.

Michael Boevink, Michael Boevink has more than 20 years management experience in the fintech and banking industry and is founder of his own investment company Boevink Group. Mr. Boevink specialises in capital raising, scaling and executing go-to-market strategies and business development growth in global markets and is engaged in companies as Raimac Financial Technology - Raimac.io - a programmable payment solution. He holds an MBA from the University of Bradford.

Mike Oaten, Mike Oaten is CEO of Tikos Technologies Limited. Based in the UK, Tikos is developing tools for building trustworthy AI with an focus on explaining outputs from deep-learning 'black-box' models.

Mitko Karushkov, Mitko Karuskov has been providing legal, regulatory, compliance, transactional and business solutions to international companies for more than 20 years now. Focused on enterprise companies and their strategic (or daily) operations, Mitko has solved matters related to the digital, tech or electronic assets of such businesses. Active and involved also in bridging between traditional and technology markets, including to the application of the EU DSA, DMA, AI and other regulations. Media, Telecoms, IPRs, Corporate, M&As are also part of the service portfolio of Mitko. For further information: www.karushkov.com.

Nadine Soyez, Nadine Soyez has been a management consultant since 2005, specializing in the intersection of business, technology, and AI-driven transformation. With extensive experience working on diverse projects across enterprises, mid-sized companies, and leading management and IT consultancies, she is passionate about how technology is reshaping the way we work, think, and grow. Her mission is to help individuals and organizations harness AI and digital technologies to drive innovation, efficiency, and long-term success. With expertise in business strategy and technology, she bridges the gap between these domains to create practical, results-oriented strategies. Clients appreciate her hands-on approach in leveraging AI for smarter decisions, digital collaboration, and business process optimization.

Prof. Ingrid Vasiliu-Feltes, MD EMBA, Prof. Dr. Ingrid Vasiliu-Feltes is a visionary leader operating at the intersection of academia, business, government and not-for-profit sectors, recognized globally for her deep tech diplomacy and digital ethics efforts. With over two decades of executive experience, she has held numerous high-impact leadership roles and has extensive complex system integration expertise, driving the development of responsible, inclusive, diverse, sustainable AI, blockchain and other deep tech innovation ecosystems at a regional, national or international level. Her unique background positions her as a thought leader on how emerging or frontier technologies are posing unique ethical challenges and are reshaping law, regulatory frameworks, corporate governance, risk management, compliance and enterprise digital strategy. She is an alumna of MIT, Harvard, Stanford, Columbia University, and University of Miami's Herbert Business School. She is a Lean Six Sigma Master Black Belt, holding executive certifications in AI, Blockchain, Finance, Mediation, Tech Diplomacy, Human Rights, and Ethics. She has served as an expert advisor to numerous Fortune 100 and 500 companies, US DOD, IEEE, NIST, and EU, UN or G20-affiliated organizations, guiding them on strategic decisions around digital transformation, digital risk governance, digital trust, and digital cyber-ethics orchestration."

Vibhav Mithal, Vibhav Mithal is an Associate Partner at Anand and Anand and is practicing as an intellectual property litigator for over 8 years. Vibhav has been a part of many path breaking litigations such as the Aloys Wobben dispute (Supreme Court, 2014); Roche v. Cipla (Delhi High Court, 2015); Shree Nath v. ABD (Delhi High Court, 2015); Monsanto dispute (Supreme Court, 2019), Ferid Allani (Delhi High Court, 2019 & Intellectual Property Appellate Board, 2020) and Armasuisse (Delhi High Court, 2023). Vibhav regularly contributes to leading IPR publications such as Managing Intellectual Property magazine, Computer and Telecommunications Law Review, Asia Business Law Journal, the Intellectual Property Law Review and Patent Litigation Review and has also co-authored the India Chapter in Global Patent Litigation (3rd Edition, 2019) published by Bloomberg Law. Vibhav has also been recognized by Managing Intellectual Property as a Rising Star, 2022 and 2023.

Zafar Imran, Zafar Imran has a track record of successfully managing Value Engineering teams & Cloud Adoption for customers. Successful track record of helping 500+ customers in the EMEA, ECEMEA & APAC region towards their Digital Transformation projects across Public, MRD, CMUT, Oil & Gas, E& C, FSI industries. Zafar Imran has 30+ years of experience with global organizations and MNCs in APAC, EMEA and ECEMEA.

Doug Hohulin,
Business Associate, AI & Partners

Vibhav Mithal,
Associate Partner, Anand and Anand

Michael Boevink,
Founder, Boevink Group

Anandaday Misshra,
Founder & Managing Partner, AMLEGALS

Martin Heitmann,
Guiding Teams to Success

Colin Crone,
Director, ISOCO(UK) Ltd

Charles Kerrigan,
Partner, CMS UK

Prof. Ingrid Vasiliu-Feltes, MD EMBA,
Founder and CEO, Institute for Science, Entrepreneurship and Investments

Harvey Castro, MD, MBA,
AI Advisor & Healthcare for Singapore Government

Lisa Ventura,
Founder, Cyber Security Unity

Daniel Ballin,
CEO, Ideas Crucible

Patrick Orsos,
Founder, mgoIT

Dr. Benedikt Kohn,
Tech Attorney | AI Regulation | IT & Data, Taylor Wessing

Nadine Soyez,
Founding Partner & AI Accelerator, Designing AI Heroes

Ginés Sánchez Navarro,
CEO, Qsimov

Zafar Imran,
CEO, Sabza

Maury Shenk,
Founder & CEO, LearnerShape

Arjan ten Buuren,
Managing Partner, Xablu Venture Studio

Ina Schöne,
Founder,
Data Privacy and AI

Hande Ocak Başev,
President,
WSI London

Guy Parsons,
Clinical AI Expert and Digital Health Advisor

Mike Oaten,
CEO, Tikos

Mitko Karushkov,
Founder,
Karushkov Legal Solutions

Arno Debelle,
Attorney-at-Law | M&A |
AI Governance & Compliance Expert | Data Scientist

David Kohnstamm,
Co-Founder and Chief Sustainability Officer, Leafcloud

Important notice

This document has been prepared by AI & Partners B.V. for the sole purpose of enabling the parties to whom it is addressed to evaluate the capabilities of AI & Partners B.V. to supply the proposed services.

Other than as stated below, this document and its contents are confidential and prepared solely for your information, and may not be reproduced, redistributed or passed on to any other person in whole or in part. If this document contains details of an arrangement that could result in a tax or National Insurance saving, no such conditions of confidentiality apply to the details of that arrangement (for example, for the purpose of discussion with tax authorities). No other party is entitled to rely on this document for any purpose whatsoever and we accept no liability to any other party who is shown or obtains access to this document.

This document is not an offer and is not intended to be contractually binding. Should this proposal be acceptable to you, and following the conclusion of our internal acceptance procedures, we would be pleased to discuss terms and conditions with you prior to our appointment. Images used throughout the document have either been produced in-house or sourced from publicly available sources (see **References** for details).

AI & Partners B.V. is the Dutch headquarters of AI & Partners, a global professional services firm. Please see https://www.ai-and-partners.com/ to learn more about us.

Designed and produced by AI & Partners B.V.