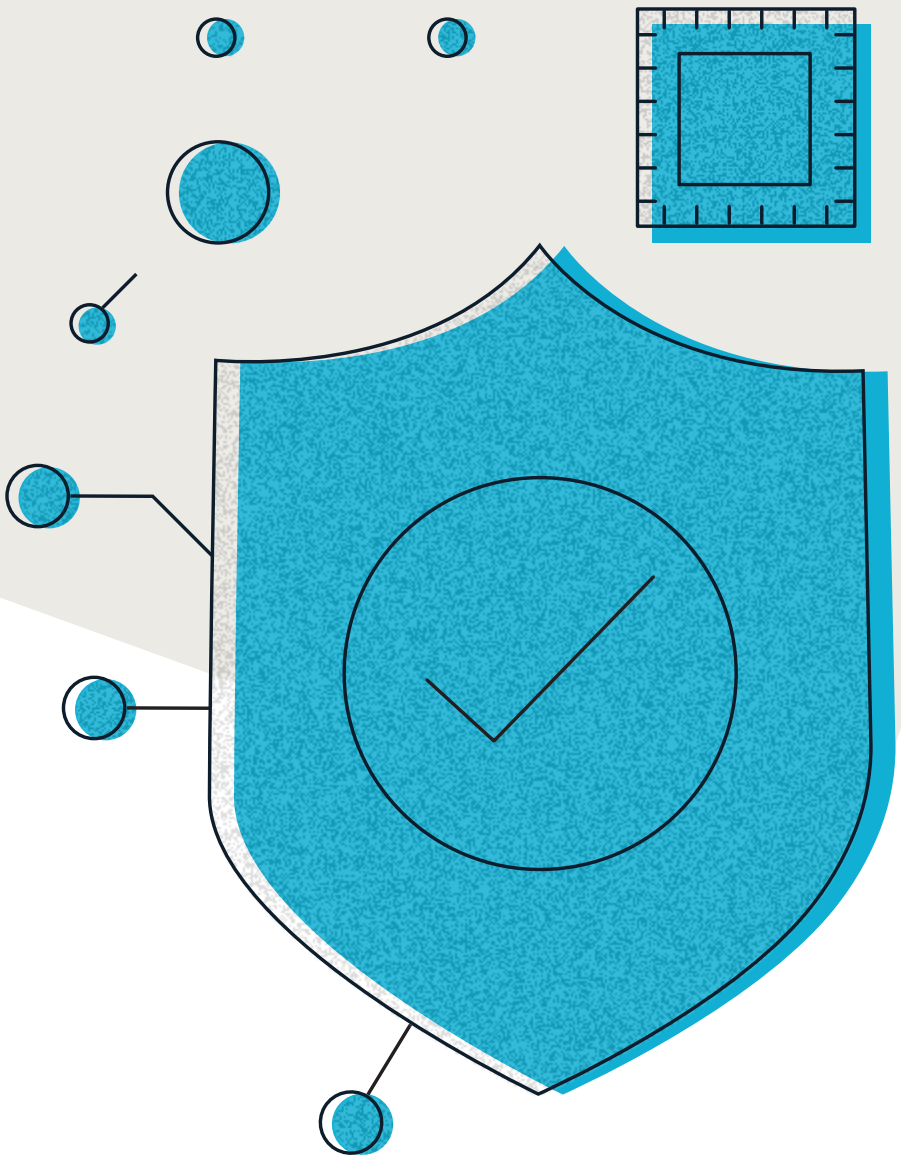
Innovation Briefing

AI assurance: protecting next-gen business innovation



cambridgeconsultants
Part of Capgemini Invent

CC in three
Your key notes to take away

1. AI will transform business and society – but it must be developed in a responsible way that maximises its potential and minimises its risks

2. AI assurance will ensure that your solutions are reliable, trustworthy, safe and secure – improving innovation and helping with future regulations and standards compliance

3. To succeed, you must focus on the entire AI technology ecosystem – and across the lifecycle of a product, service or solution. A holistic approach is key

Introduction

The transformative capabilities of AI offer enormous promise. Huge gains have already been demonstrated in areas as diverse as medicine, biotechnology, manufacturing, telecommunications, agriculture and transport. But many innovators are urging a more responsible approach to development because AI has the potential to become extremely powerful and transformative – a situation which could provoke significant societal and economic impacts.

These effects could be unintentionally negative, particularly in areas where ethical questions need to be careful considered, such as autonomous security or decision-making systems. Impacts could also be deliberately malicious, from autonomous weaponry to cyber-attacks. Fears persist too that AI, especially when combined with automation, could replace human workers, leading to job losses in certain industries and increased inequality.

So how to move forward? Thought leaders are calling for AI to be developed in a responsible and beneficial way that maximises potential while minimising risks. Nations and standards bodies are developing and evolving regulations and legislation to ensure that caution and considered action is our default approach.

Nonetheless, many argue that these frameworks and processes do not go far enough or quickly enough given the rate of evolution of AI technology. But the reality is that this technology cannot be ignored. Those who successfully navigate AI implementation AI into their business in a responsible, transparent and effective way will lead their sectors in the future.

Where does this leave you as a business leader? In this Innovation Briefing, we unpack the crucial points you need to consider when implementing AI. The objective is to maximise impact for your organisation while futureproofing it against the exponential changes in technology, market and societal demand, and regulatory requirements.

At Cambridge Consultants, our comprehensive approach to innovation ensures that we remain at the forefront of AI. And our data driven-methodology grounds us in reality, to help us better serve industry and society on a global scale. We're well placed to support your ambitions, so let's get down to business with the proactive steps you can take to get ahead of the curve…

Defining AI assurance and its commercial and societal value

The benefits

AI assurance is the process of ensuring that systems and solutions are reliable, trustworthy, safe and secure. Positioned at the intersection between AI trustworthiness and user trust, it involves a range of activities that include designing, developing, testing, validating, and monitoring AI systems. The purpose is to meet the intended requirements and comply with relevant regulations and standards, while remaining transparent and explainable to users and stakeholders.

Assurance is of significant value to business because it:

- Enhances the trust and confidence of users and stakeholders in AI solutions, which can bring about increased adoption, customer satisfaction and brand reputation
- Addresses ethical and social concerns related to AI, such as fairness, privacy and security, which can promote social responsibility and avoid negative impacts on society
- Accelerates innovation and enhances human creativity. When properly designed and developed, AI systems can provide businesses and organisations with new insights, ideas and opportunities previously inaccessible or overlooked. They can help businesses identify patterns, relevant correlations, causations and relationships in large data sets, generate new hypotheses and predictions and suggest novel solutions to complex problems
- Improves the reliability and performance of AI systems, leading to better decision making, increased efficiency and reduced operational costs

The result is enhanced trust in intended performance that builds the confidence of users and stakeholders. Ecosystems are compliant with relevant laws and regulations and are able to address ethical and social concerns related to AI. Businesses and organisations are then empowered to create new products and services, streamline processes, optimise operations and identify new opportunities for growth and expansion – which can be implemented sustainably.

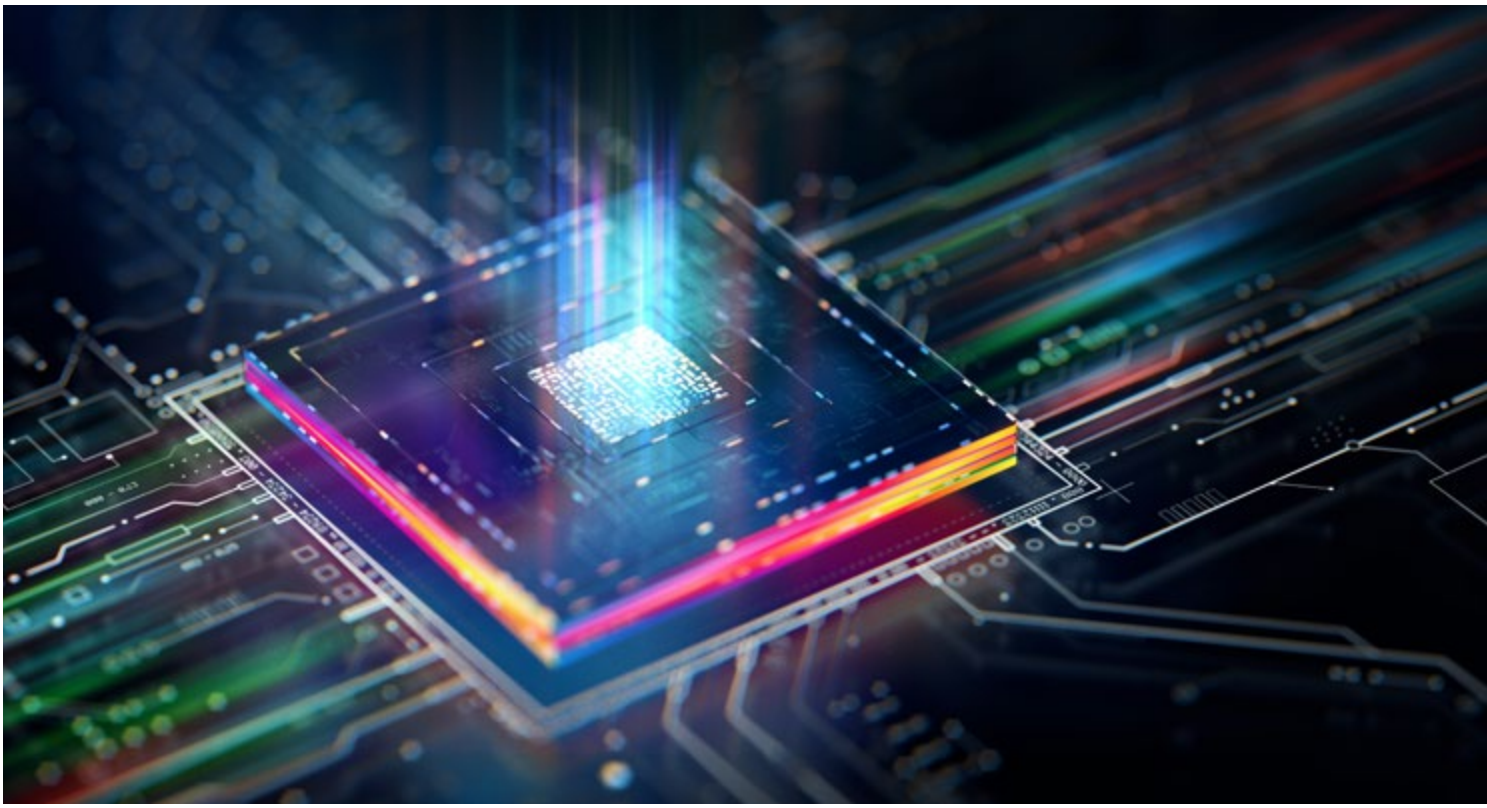
The societal impact

The societal value and potential impact of AI is very significant, particularly when it is deployed in sensitive or critical domains. AI assurance can promote the adoption and diffusion of AI across different sectors and industries while identifying and mitigating any societal risks. There is currently a boost at country level, signified by governmental AI strategies being published around the world.

In the United States, the blueprint for the AI Bill of Rights was published in October 2022 and the NIST (National Institute of Standards and Technology) Risk Assessment Framework 2.0 followed in January 2023. Meanwhile, the EU is working on a comprehensive directive based on risk-assessment of AI systems which is anticipated to be approved soon. Similar initiatives and frameworks are being developed around the world. International bodies such as the OECD has created a framework to assess trustworthiness of AI systems, while the UN has released its guidance to ethical AI.

Industries such as health, automotive, aviation and manufacturing, to name a few, are trying to navigate the current technical advances versus assessment of risks. The aim is to be sufficiently prepared to assure the safe and secure incorporation of AI into complex systems and environments. Finally, international standard institutions such as ISO and IEEE are working on several technical reports to address assurance issues.

All this activity indicates the criticality of the challenge and the impact on society at all levels. Business, industry and organisations are under pressure to understand and mitigate risks while seizing transformative technological developments. Risks likely to arise from AI must be identified and tools must be developed to determine the necessary mitigations. This is why AI Assurance is so crucial. It is not an option, businesses that fail to respond will be rapidly 'out innovated' by those that do.



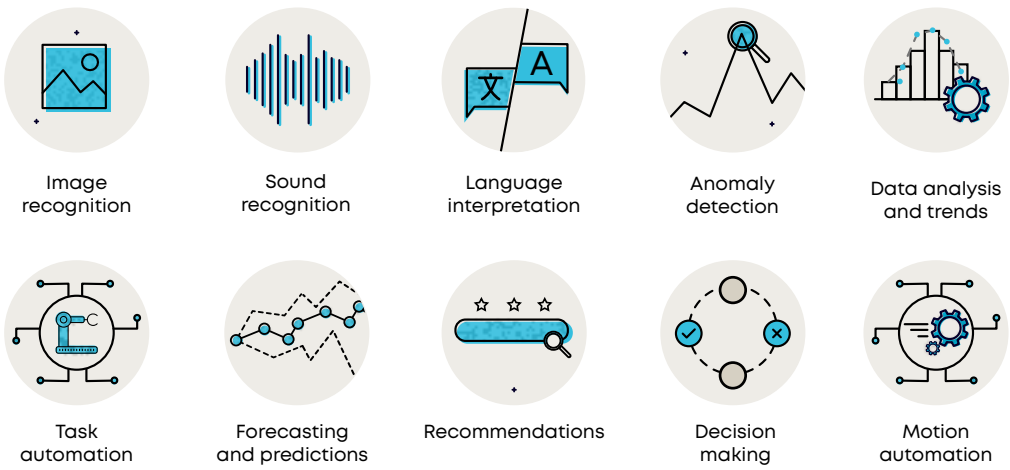
Industry 5.0 and augmenting humanity

In general, AI's purpose is to replicate different characteristics of human perception, analysis and decision making in order to process vast amounts of data much more quickly and efficiently. It also analyses these data, finds more insights, produces forecasts, recommendations and decisions, and achieves the automation and autonomy of tasks. These AI specialist areas are used in various industries and in different applications.

For example, there are AI solutions that are focused on interpreting images and detecting anomalies (medical x-rays, for instance), natural language processing (voice assistants, automatic translations, automatic text generation), physical functionality and motion (autonomous robotics, cobots, autonomous vehicles) and predictivity and forecasting (recommendation engines, forecasting future behaviours, decision making, targeted marketing/ messaging). Additionally, different algorithms developed and used in each of these areas can combine to create innovative solutions and complex systems.

MOST COMMON APPLICATIONS OF AI

through a technological solution-focused lens



In this sense, AI is also a vital component of industry 5.0 – a concept that builds on the principles of industry 4.0. the latter is focused and powered by smart technologies whereas industry 5.0 emphasises the role of human skills, creativity and social responsibility in the design, development and deployment of advanced technologies such as AI. It envisions a future where humans and machines work together in a collaborative and complementary way to create value for society and the environment and promote the wellbeing of individuals and communities. Therefore, AI assurance becomes essential within industry 5.0.

In order to be transformative, technology needs to benefit humans, society and the environment. Assurance creates trust and understanding of AI, analyses its interactions within complex systems and with humans, identifying risks at different levels and putting mitigations in place within an integrated process.

In addition, AI assurance can play a key role in realising the vision of industry 5.0 by ensuring that AI systems are designed and developed in a way that aligns with human values, ethics and aspirations – and that enable humans to augment and enhance their capabilities, rather than replacing or undermining them.

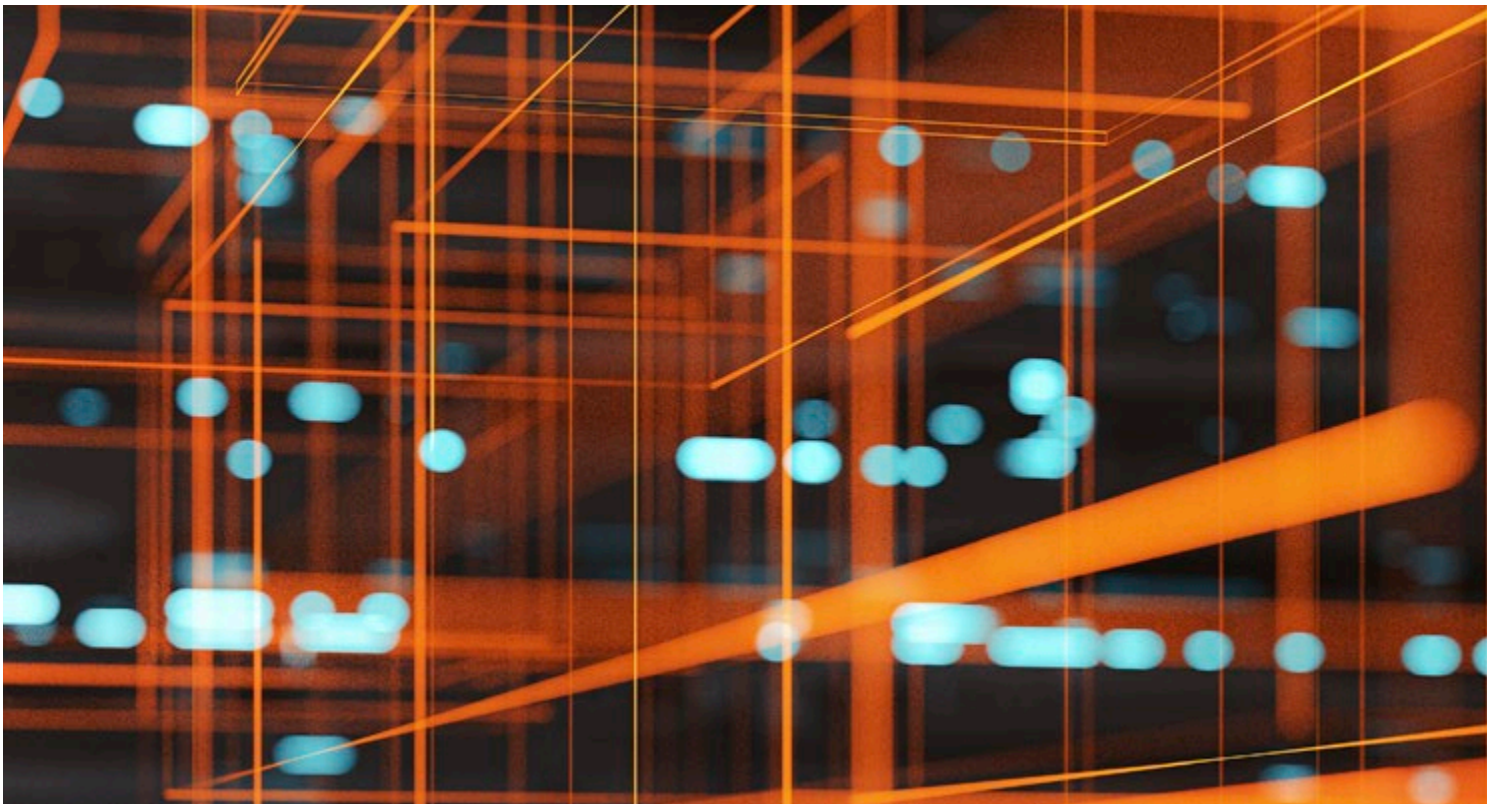
The challenges of implementing AI assurance

For the effective and impactful implementation of AI assurance, a number of challenges must be addressed. Chief among them is the need for businesses to focus on the nature and characteristics of the entire AI-based technology ecosystem across the whole lifecycle of a product, service or solution.

Failing to implement any form of AI assurance has resulted in some high-profile cases where organisations are left dealing with reputational fallout long after they have 'fixed their issues'. Worse still, an error by one organisation can have repercussions across all sectors. This is because a 'failure of AI' becomes a catch-all statement for AI shortfall.

A couple of examples from recent years highlight this negative impact on perception. In 2018, Amazon developed an AI system to screen job applicants, but it was later discovered that the system was biased against women. The system was trained on résumés submitted to Amazon over a 10-year period, which were predominantly from males. As a result, the AI system learned to favour male candidates and penalise female ones.

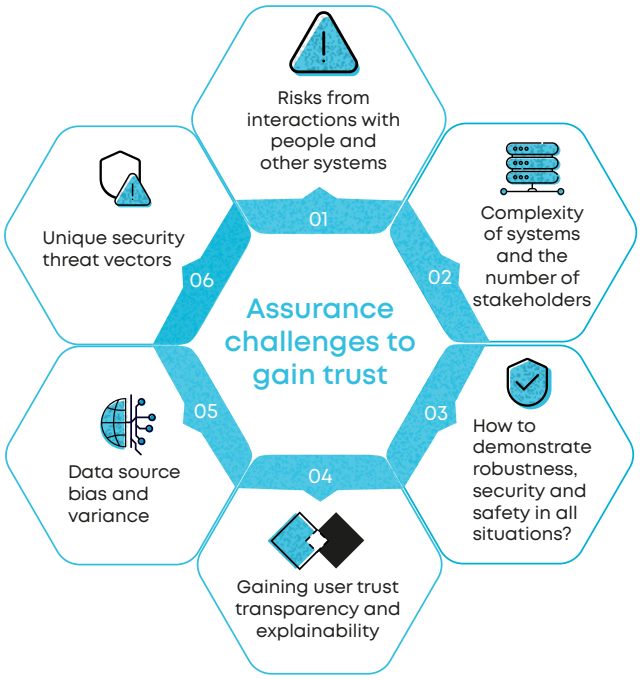
This AI bias in hiring tools has become such an issue that New York City Council passed a bill that regulates employers and employment agencies' use of 'automated employment decision tools' in making employment decisions. But the stakes are even higher. Self-driving cars have the potential to revolutionise transportation, but they must be safe and reliable. Also in 2018, clearly a pivotal year for scale implementation of AI, a self-driving Uber car struck and killed a pedestrian in Arizona. An investigation found that the car's sensors had detected the pedestrian, but the AI system had failed to apply the brakes. More recently, in 2022, Telsa recalled around 50,000 vehicles after the AI failed to stop at signs.



The challenges of implementing AI assurance

Act now to develop best practice

There is, however, a momentum within many industries to incorporate automation in their flowlines, products and services by embracing AI. Traditionally, each type of industry defines how a new technological solution might comply with existing standards, regulations or certifications. But when considering AI systems, current industry regulations, standards and certifications are simply inadequate or unable to cope with the new characteristics and risks. Upgraded standards and guidelines are arriving but very slowly, so organisations need to act now to develop best practice, expert knowledge and AI assurance techniques to ensure that their adoption of AI is done in a responsible manner.



Here are just some of the new challenges brought about by AI systems:

- They are highly dependent on data and/or knowledge inputs for training, testing and performance
- They are probabilistic in nature as the rules created by the algorithm can sometimes change with minimal modification in inputs which create great impact on outputs
- Their performance might deteriorate with time. There is a possible data and model behaviour drift as data and deployment environments change
- Lack of transparency and model interpretability – and as consequence a lack of trust from operators and users
- The added complexity of interactions with other solution components, with other AI agents and with external factors
- Understanding human-machine interactions, transparency, explainability and the implications for building trust
- The enormous potential for AI automatic and autonomous outputs and decisions to positively or negatively impact a particular task or process, individuals, society and environment

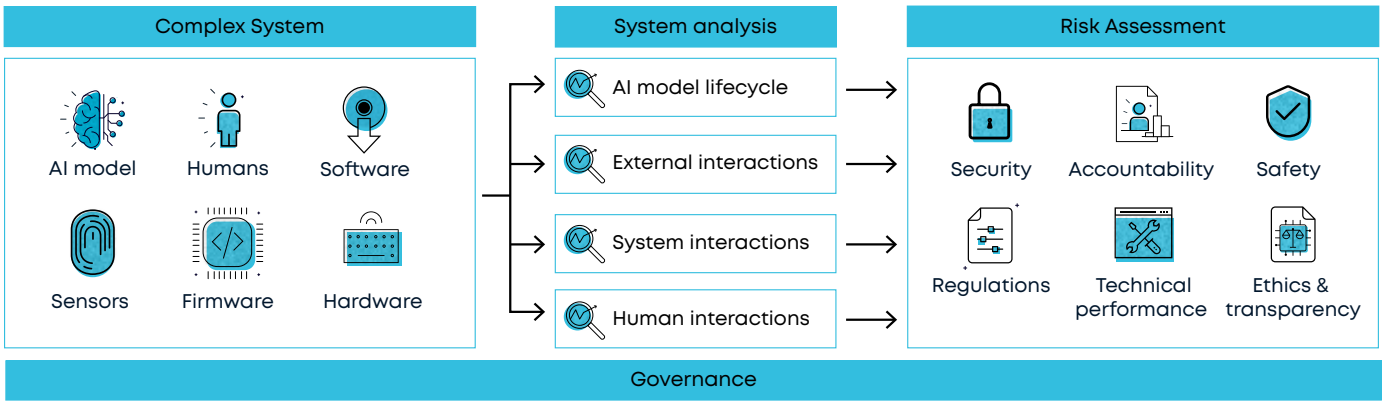
A well-defined and focused process to identify, measure and manage these challenges is crucial. At first sight, many companies might feel so overwhelmed that they concentrate on the most obvious or pressing challenge (such as data quality) and leave the others to be dealt with later. This is not a very efficient way to address AI risks, as challenges that might have been easy to solve early in the design stage are trickier later in the process. The aim of AI assurance processes and tools is to address the complexity of all the above challenges over time and in a comprehensive and inclusive way.

AI systems and their assurance process

As we have already mentioned, the new characteristics and challenges introduced by AI mean that novel risks need to be identified and managed. Some AI risk assessment processes can be built upon established ones without reinventing the wheel, but they would need to be updated and upgraded to cater for specific risks.

How to break down complexity

For the early identification of any challenges or risks, it is crucial to break down the complexity, interactions and interdependencies of AI within and between complex systems during automation roadmaps.



An AI model doesn't normally exist in isolation, it is usually part of a more complex system formed by elements such as software, hardware, sensors and humans, with multiple interactions in between. So, an important first step is to understand the characterisation of the complex system, the AI lifecycle itself and the types of interactions that exist within and outside the system. Understanding all these aspects will help to identify areas and touch points to analyse further. Interactions include:

- Human-machine interactions. The human is at the core of any AI solution and any involvement in the process, oversight and impact that the AI might have are a top priority
- External interactions with the environment in which the solution is operating
- Internal interactions with other components or other AI(s) within the solution

Once we understand the complexity and the multiple interactions of the system and the AI within, the potential risks introduced by the AI component can be mapped. This should be done while considering different aspects of assurance – from safety, security and technical performance to regulations, ethics, sustainability and transparency (and the knock-on effects between them). Overarching all this is a need for a process of oversight and accountability. These risk assessments are based on evidence which needs to be documented and owned. In this way, responsibilities are clear and delineated, and interactions between different areas are understood.

Needless to say, not every AI assurance process will need to go deeply into all the different possible risk areas that we have illustrated. Everything depends on the complexity of the solution and the algorithm, the industry where is going to be applied, the environment that it is going to operate in and whether there is strong need for human interaction with the system and the potential impacts.

AI assurance use cases

Agritech

Within the agricultural industry there is plenty of research investment focused on how automation could help with time-consuming tasks that need intensive human effort. This in an industry that's demanding solutions – and automation is very much in line with the philosophy of industry 5.0. The aim is to automate repetitive and demanding tasks which present challenging conditions for humans, such as fruit picking or toiling in high temperatures.

Automation and autonomy could also be introduced for measuring plant needs (soil nutrients), targeting irrigation where it is needed and optimising resources to increase productivity. The automation of any these tasks means an assessment of:

- Is automation/use of AI suitable for the task? (Some agricultural tasks might not be suitable for automation, or the type of automation needed might have not been developed yet due to different constrains)
- The purpose for which machine learning or the AI solution is needed (for example, to create an autonomous robot/vehicle that needs to move between plants and pick ripe fruit in order to be more targeted, productive, sustainable and provide assistance to humans working in difficult conditions)
- Characterising the parts and flow of the solution while understanding the complexity of the system (where it sits within the different agricultural roles, tasks and flows)
- Internal interactions (between sensors, tools, measurements, data capture and hardware)
- Human-machine interactions (are humans going to work alongside the robot/vehicle?)
- Interactions with the external environment (weather, terrain, different crops, other tasks)
- Safety and security risks of the solution (where could harms occur? How does it avoid obstacles/humans? Are there any cybersecurity risks?)
- Technical performance risks (for example, is the robot/vehicle going to work in all weather conditions, does it have errors when identifying fruit/vegetables that are ready to be picked?)
- Accountability, ethics and transparency of the solution (how to explain the function of the robot/vehicle to the end user/bot co-workers, how to create trust and how to manage change)

AI assurance use cases

Telecommunications

The telecoms industry is investing to create innovative solutions that optimise performance, increase robustness and assure the sustainability of networks. These solutions can often involve machine learning or AI algorithms to identify trends and, for example, optimise the network to target areas where more or less power is needed at any given time.

In order to achieve this, there is a need for assurance and risks to be identified as early as possible, as getting the solution and automation right are normally quite consequential. This impacts many other systems that are dependent on it. Also, such projects are usually costly so any savings in time and resources from the start would make a huge difference to the final project balance sheet.

AI and machine learning solutions can help the industry be more resilient and sustainable, preparing it for future developments and engaging with industry 5.0. For this, we would always recommend:

- An initial assessment of the project. The goal needs to be clear (optimisation of a network in order to save energy, for example)
- Initial risks identification and prioritisation (breaking down the complexity of systems and processes as explained earlier)
- An initial assurance strategy with main risks to focus on during the project (how can we monitor the AI performance and reduce the risk of data/concept drift?)
- Detailed risks identification (technical performance, safety, security, sustainability and so on) and creating mitigations during the design and development processes (for example, how is the AI susceptible to new attacks and what are the implications for the network using the AI for optimising resource?)
- Validation, integration and operational deployment, where tests are performed to address AI risks, AI interactions, simulations of different scenarios and edge cases, stress testing and demonstration of safety, security and performance.
- Monitoring, fail safe/backups put in place for edge cases (if the AI failed to identify bandwidth needs and as a consequence a shutdown might occur, the mitigation could be a human on the loop alerted to help avoid the situation)
- Demonstration of overall transparency, governance and accountability measures

AI assurance use cases

Warehousing, manufacturing. lab cobots

There is no clearer case for human centric solutions than in an operating environment where robots and humans need to work closely together. The interactions might be very variable between them, but as automation gets more advanced and complex, cobots will need to work tightly and cooperate with humans in very close physical environments. New ways of working might be defined and humans might develop new skills, within the vision of industry 5.0.

For this we will need to assess and understand in depth the following:

- A holistic view of cobots tasks and system analysis
- Identification of interactions (internal and external)
- Breaking down complexity of systems and flows
- Human-machine interactions (tasks, dependencies, operational flows within the relevant context such as manufacturing, lab, warehousing)
- Technical performance assessment (tests, simulations, of cobots and humans working together within synthetic environments where edge cases can be tested and safety and security scenarios modelled to identify areas for mitigations)
- Mapping safety and security risks for humans within all the flow and interactions, identifications, and mitigations
- Integration of cobots with human practices, creating human interfaces and applying human factors (creating and sustaining trust)
- Transparency and accountability (how to explain the role of an AI to end users, how to understand and interact with the tasks performed by the cobot, how to monitor performance)
- Understanding human cooperation, human augmentation by AI assistants, how humans can be trained in new skills/tasks working with cobots, human-machine teaming and definitions in different contexts.

AI assurance use cases

The key for a successful automation to autonomy journey is an early risk assessment that considers the purpose, design, context and interactions of AI. There should also be a continuous process during the design, development and validation of complex systems. For example, if we identify risks related to specific AI cyber-attacks such as data poisoning, the early stage design can include measures to prevent and mitigate these risks.

If we identify risks related to the human-machine interaction related to safety issues, we can start looking into situations when the AI decision drives a component to cause harm, or bias mitigation and identification on data and algorithms outputs that might cause a negative impact.

A phased identification of these risks should be the optimal flowline to follow. For those already down the path of development or validation, it is possible to pick up assurance assessment process, although the risks might mean going back into design or planning in some cases. Therefore, early intervention would save time and build trust between stakeholders as the system is developing with a full risk assured and evidence-based deployment phase.

It is really crucial to incorporate this phased approach into a company's current research and development flowlines. In this way, AI assurance is embedded within working practices and developed and assessed at the same time as technical design and development – creating 'AI assurance by design'.

	Phase 0: Concept	Phase 1a: Design	Phase 1b: Development	Phase 2: Validation, integration and deployment	Phase 3: Operational monitoring
Focus for Assurance	Development of the plan	Assurance activity informed by design discovery	Assurance activity focused by initial plan	Assurance activity informed by validation	Final assurance activity
Supporting	Solution characterisation	Design plan and execution	Solution development roadmap and execution	Solution validation Deployment roadmap and execution	Monitoring plan
Activity	Initial risk identification and priortisation	Risk identification Process evidence Update assurance plan	Risk update Testing Update assurance plan	Risk updates Testing Evidence compilation	Risk updates Monitoring and alerting plan
Outcome	Initial Assurance Plan	Assurance evidence of design and process	Assurance evidence from development testing	Assurance evidence validation and deployment	Assurance evidence monitoring

Putting it all together. A holistic AI assurance approach

A holistic approach to AI assurance

FOCUSED

Driven by an early Assurance Assessment and prioritisation plan

INTEGRATED

Process that avoids the stovepipes and thus reduces costs

COST EFFECTIVE

Testing includes Synthetic environments and data

EXPLAINABLE

Considering all stakeholders and end-user explainability to build trust

HOLISTIC

Covering safety, security, ethics, HMI, regulations, technical performance

DEALING WITH COMPLEXITY

Identifying interactions and interoperability with other systems

HUMAN - MACHINE TEAMING

Adopting a Human-Machine teaming approach to optimise design

As we have seen previously, a risk-focused assurance process needs to be holistic and cover potential risk areas such as cybersecurity, safety, ethics, regulations, technical performance and human-machine interactions. It also needs to deal with complexity, interactions and interoperability with complex systems in order to assess the new challenges added by AI. And for it to succeed, this process needs to be integrated within current flowlines, while identifying and prioritising risks.

This integration within flowlines and processes also apply to Machine Learning Operation Operations (MLOps) and Artificial Intelligence Operations (AIOps). Assurance processes need to be closely linked to the lifecycle of machine learning and AI algorithms, assessing the risks and integrating mitigations as the algorithms and models are developed so that it is easier to validate, assure and integrate them within operational environments. This in turn means saving time and resources that do not need to reassess or modify models at a later stage.

Another very significant aspect of a risk assurance process is that the AI and its interactions are understood by key stakeholders and users. It is therefore necessary to have explainability at the right level of expertise. A key for this is that the outputs of the AI are interpretable (as much as possible) and traceable though the model lifecycle. In this way, these can be understood or challenged by relevant stakeholders and users which in turn will help to build trustworthiness in the system.

An innovative approach to establishing whether a migration plan would be effective is to use the testing technology of synthetic environments, as suggested in the previous use cases. Within them, we can create unique scenarios for AI risk identification, management and mitigation with cost-effective methods to create robust mitigation plans.

Finally, and at the centre of all these processes, are human beings. AI assurance allows in-depth thinking about impacts on humans, human-machine interactions, human-machine flowlines, human-machine collaboration in different settings and an assessment of whether the AI, for example, will augment human capabilities or change the human role.

What next?

Very few companies have managed to develop and integrate the AI assurance processes we've described in this Innovation Briefing. But those that have are the leaders in the AI solutions market.

The reasons for such slow adoption include not understanding all the challenges of AI design and development, the complexity of systems enhanced by interactions with an AI, not defining a clear flow of targeted of actions and not creating a holistic AI assurance plan that is easy to embed within current practices.

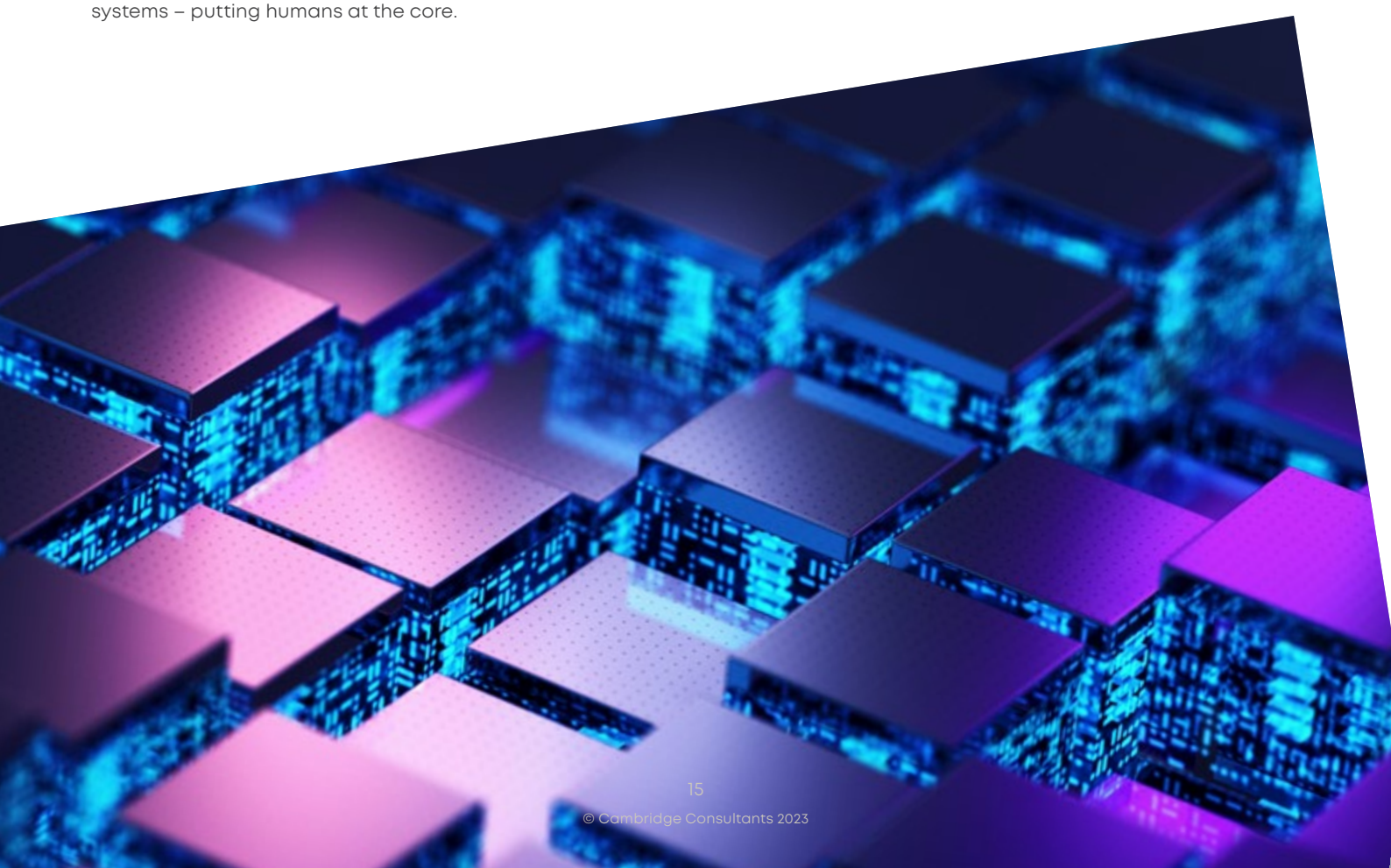
We have focused on all these issues to help companies grasp the required processes and approach AI projects confidently. It is clear that AI assurance is the new must-have for any company on the automation to autonomy journey. It is a fundamental part of the AI design, development, systems integration, deployment and monitoring processes.

This is, as we have seen, is reflected by the number of guidelines, principles, frameworks and upcoming regulations across governments, industries and regulators for AI risk assessment and AI assurance.

It is also important to get ahead and join the industry 5.0 revolution. Technology needs to work for humans, society and the environment. AI assurance puts emphasis on risks and impacts that a solution might have in very complex systems – putting humans at the core.

So, what can you do as a business? The objective should be to position yourself in the market as a responsible AI provider and demonstrate that your solutions are trustworthy and assured. Aiming for the highest level of AI maturity is the way to thrive. On that note, first considerations should include:

- Gathering internal and external stakeholders to begin conversations about the requirements and suitability of AI within the autonomy roadmap, when and where the AI should be incorporated, what tasks will be automated and what implications this will have for humans involved
- Considering any human involvement, interaction or potential impacts as a top priority
- Identifying risks that the AI might introduce in areas such as safety, security, technical performance, ethics, regulations, transparency and the links and knock-on effects between them
- Embedding the AI Assurance process within the existing flowlines and ways of working, with a governance and accountability system in place



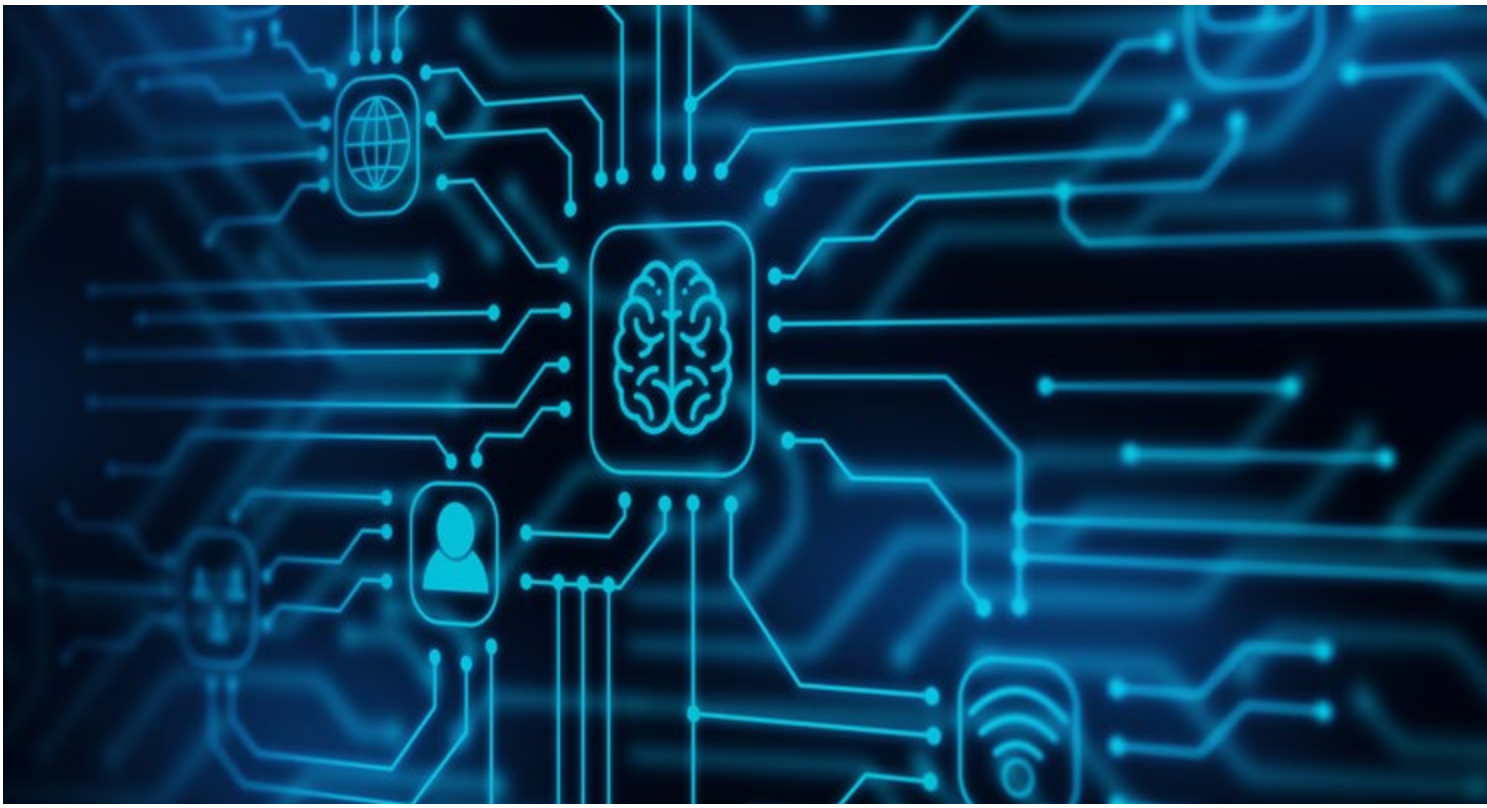
Conclusion

AI can be an incredibly powerful tool for businesses and organisations when implemented well. But the potential benefits can only be realised by taking steps to address the pain points associated with AI, such as bias, security vulnerabilities and ethical considerations.

When implemented responsibly and ethically, AI can help:

- Improve decision-making. AI can analyse vast amounts of data quickly and accurately, helping businesses make better, data-driven decisions
- Increase efficiency. AI can automate repetitive tasks, freeing up employees to focus on more complex and creative work.
- Enhance customer experiences. AI can help businesses personalise its products and services, providing customers with more relevant and customised experiences
- Support sustainability. AI can help reduce environmental impact by optimising energy usage, reducing waste and improving supply chain efficiency
- Foster diversity and inclusion. By minimising bias and promoting fairness and equality, AI can help build more diverse and inclusive workplaces

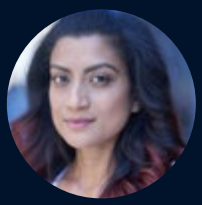
Overall, the potential benefits of AI are unparalleled, but we need to ensure that AI systems are developed and used responsibly. By implementing AI assurance and taking steps to address potential risks and challenges, we will be better placed to unlock the full potential of AI and drive innovation, growth, and social impact.



Why CC?

Cambridge Consultants (CC), part of Capgemini Invent, is a global team of 800 bright, talented people – united by the ambition to turn brilliant and radical ideas into technologies, products and services that are new to the world. We expand the boundaries of technology innovation by tackling the tough, high-risk challenges that bring sustained competitive advantage and market leadership for clients. We are trusted by some of the world's biggest brands and most ambitious start-ups to realise their critical technology-based aspirations – and we've been doing it for 60 years.

Dr Carolina Sanchez Hernandez,
Principal Assurance Technologist
carolina.sanchez@cambridgeconsultants.com



Dr Maya Dillon,
Associate Director, AI
maya.dillon@cambridgeconsultants.com

cambridge consultants
Part of Capgemini Invent

UK — USA — SINGAPORE — JAPAN

www.cambridgeconsultants.com

Lit-IB-007 v0.2 '2023_AI_Assurance_Innovation_Briefing'