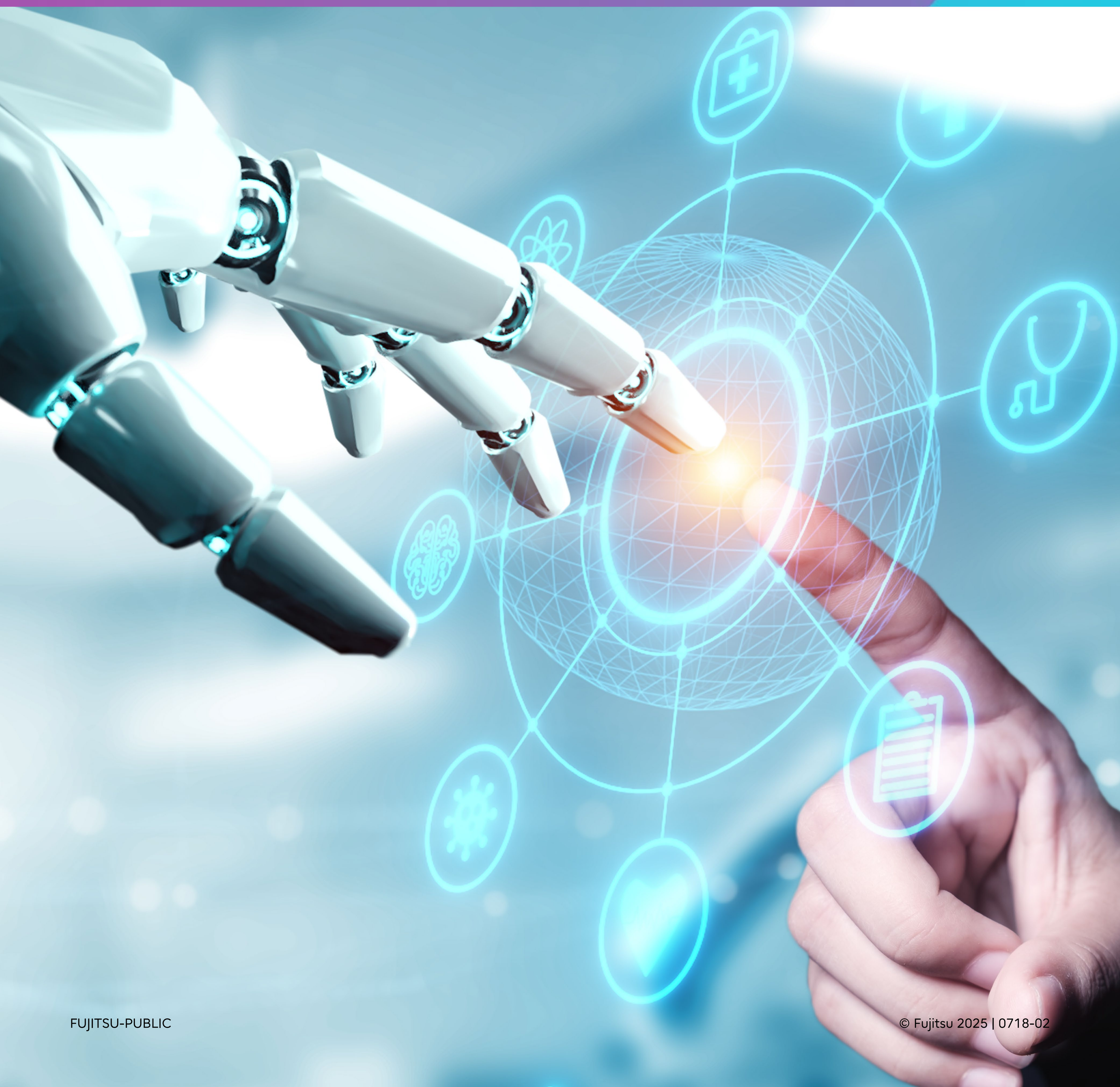
FUJITSU

Whitepaper

Responsible and ethical AI



Contents

Introduction

Companies in every industry looking to survive in the artificial intelligence (AI) age are trying to predict the future to make the right decisions about how to implement AI quickly, efficiently, cost-effectively, and, most of all, for competitive advantage. In the race to achieve that competitive advantage, there is a possibility that important ethical and governance concerns may be overlooked.

While AI holds unprecedented potential to make organisations and processes more efficient and effective, these benefits come with correspondingly unprecedented risks. Managing those risks is essential for organisations to leverage AI successfully without falling into potentially costly traps.

Legislation regulating the use of AI is on the horizon around the world, including Europe, the USA, and the Asia-Pacific region. However, companies that wait to see how this legislation evolves and is enforced are missing the bigger picture. Legal repercussions could be the least of their concerns if they engage in unethical AI practices or suffer a security breach; the reputational damage and loss of customer trust could be far more costly and destructive to long-term business value.

The potential for unethical AI use

There are a vast number of AI use cases across almost all industries. Some examples include:

Healthcare: diagnoses and treatment recommendations

AI can analyse vast amounts of medical data, including patient histories, research papers, and clinical trials, to assist in diagnoses and treatment recommendations with potentially greater accuracy than human physicians alone. AI can also facilitate early detection by identifying subtle patterns in medical imaging or test results that might be missed by human eyes, possibly catching diseases at earlier, more treatable stages. And, it can help tailor treatment plans to individual patients based on their genetic makeup, lifestyle, and other factors, likely improving outcomes.

However, one of the biggest risks is biased training data. If AI systems are trained on datasets that are not diverse or representative, they may make biased recommendations that could disadvantage certain demographic groups. There is also the risk that healthcare providers might become too dependent on AI, possibly overlooking important clinical judgments that require human intuition and experience. Privacy concerns are valid as well; the use of AI in healthcare necessitates access to large amounts of sensitive patient data, raising issues of data security and patient privacy.

Insurance: premiums and risk assessments based on socio-economic factors

AI can analyse complex patterns in data to provide more precise risk assessments, potentially leading to fairer pricing for many customers. AI algorithms can also identify patterns indicative of insurance fraud effectively to possibly reduce costs for insurers and honest policyholders. At the same time, AI could facilitate the creation of highly customised insurance policies tailored to individual needs and risk profiles.

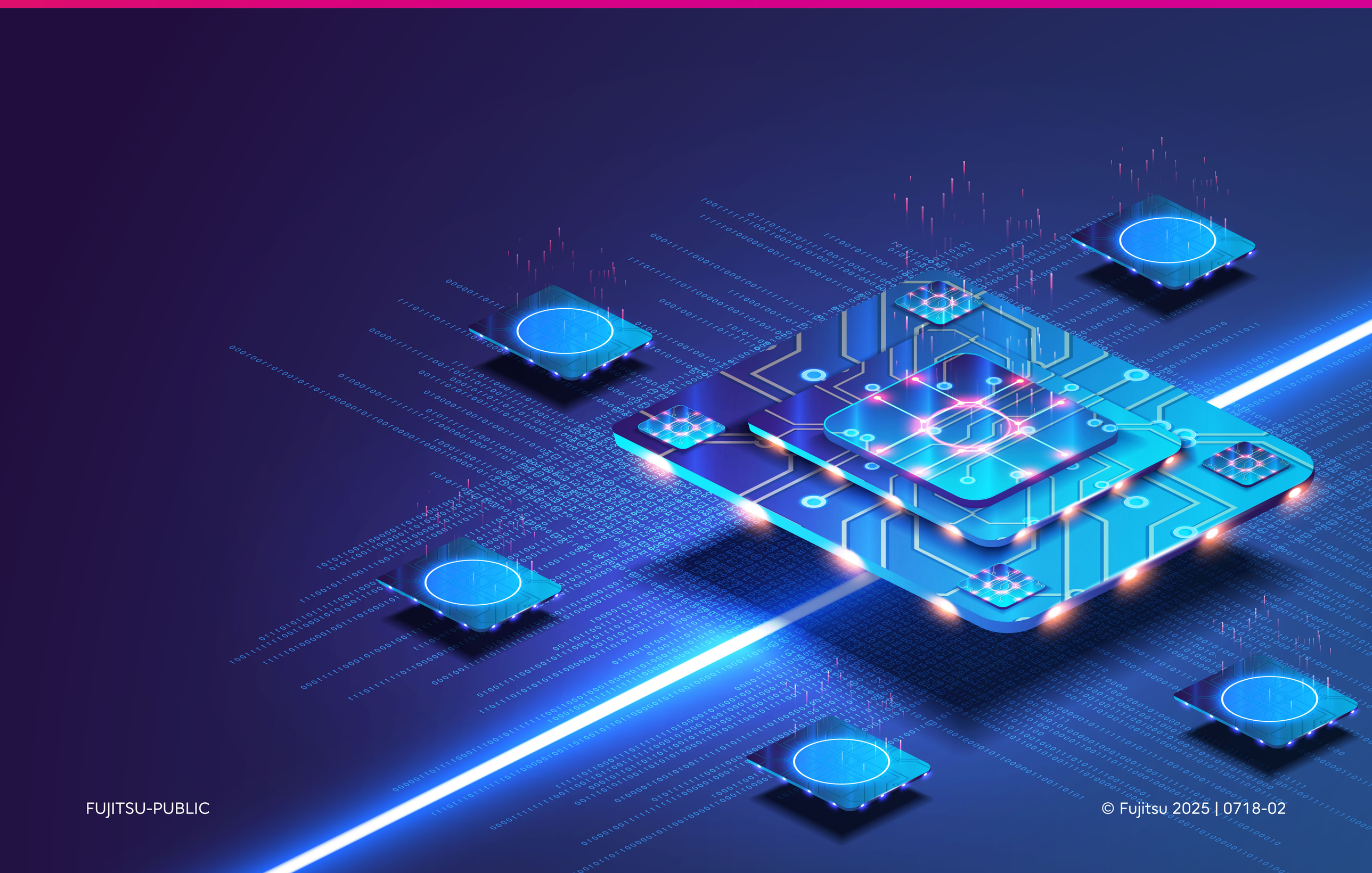
Risks include discrimination concerns. For example, if AI systems consider socio-economic factors in their assessments, there's a risk of perpetuating or exacerbating existing societal inequalities, potentially leading to higher premiums for already disadvantaged groups. Additionally, the lack of transparency can be an issue, as the complexity of AI algorithms can make it difficult to explain how certain decisions are made, which could be problematic in disputes or regulatory compliance. The use of extensive personal data for risk assessment also raises questions about data privacy and the ethical use of information.

Banking and finance: loan decisions, interest rates, and investment decisions for superannuation

AI can process loan applications and make investment decisions much faster than traditional methods, improving efficiency and customer experience. AI can also analyse a wider range of factors when making lending decisions, potentially opening up credit opportunities for those who might be overlooked by traditional methods. For superannuation, AI can continuously analyse market trends and adjust investment strategies in real time, possibly leading to better returns for retirees.

However, there is also a high potential for bias. If not carefully designed and monitored, AI systems could perpetuate existing biases in lending practices, potentially discriminating against certain groups. Complexity is also an issue in banking and finance, where it becomes difficult to explain why a loan was denied or how an investment decision was made, which could be problematic for regulatory compliance and customer trust. These issues could be exacerbated if many financial institutions rely on similar AI models, as this could lead to herd behaviour in financial markets, possibly creating market volatility or systemic risks.

With so much opportunity for AI to replace human-centred decision-making, organisations must act now to mitigate these risks, protect their reputations, and benefit from AI without stepping on legal and ethical landmines. This will require organisations to understand and adopt ethical AI principles.



General principles for ethical AI

AI solutions developed or used by an organisation should adhere to a foundation of ethical AI principles. These principles should be tailored to the organisation's specific context and challenges, and should be clear, actionable, and aligned with industry best practices.

The relevant ethical frameworks will vary from business to business; however, some ethical considerations are universal. These include:

Fairness and non-discrimination

It's essential for datasets used in training AI models to be diverse and inclusive to prevent potential bias against any demographic group. Organisations should also implement rigorous testing procedures to identify and address biases in AI models, including regular audits of model outputs across different demographic groups. At the same time, they should develop and apply appropriate fairness metrics to evaluate AI system performance, considering various definitions of fairness and their applicability to specific contexts.

Transparency and explainability

Modern AI models are extremely complex and, in many cases, even their developers struggle to explain how they come up with a specific decision or prediction. For business cases where explainability is important, this can limit the types of models and mathematical techniques that can be used to those where a clear reason for decisions can be provided. This can be important for cases such as in banking when determining whether to approve or decline a loan.

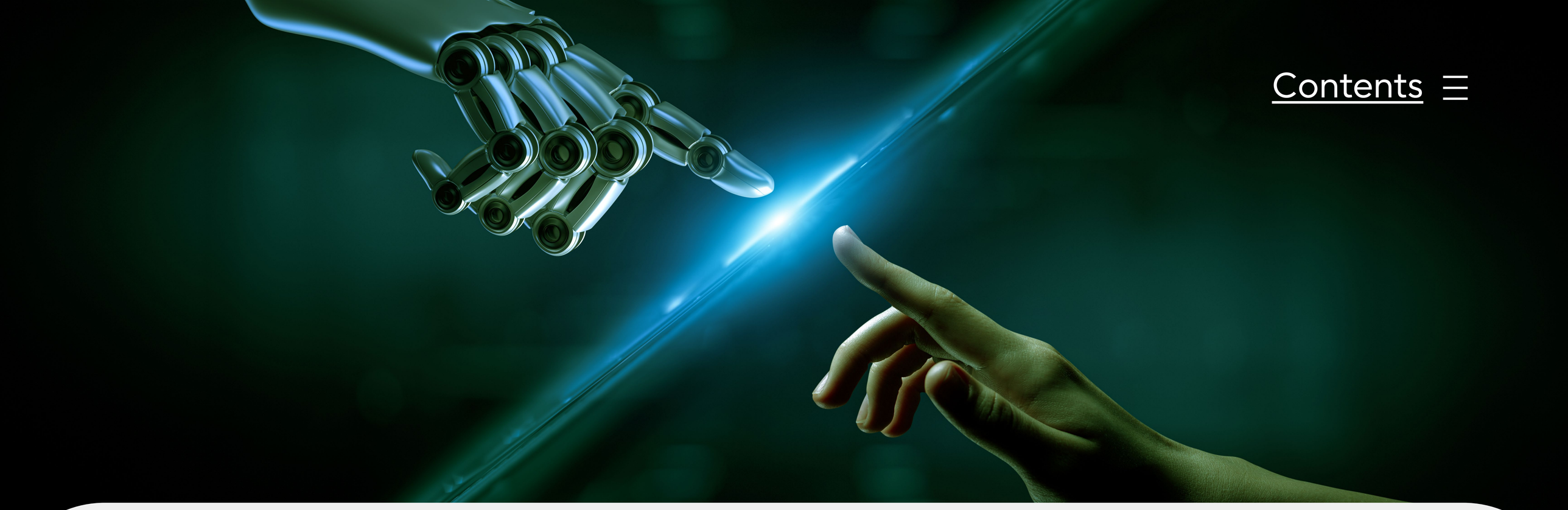
Organisations also need to be transparent with their stakeholders on the use of AI. Where a decision has been made or is impacted by an AI solution, this should be communicated to stakeholders clearly. This includes providing information about the purpose of the AI, the types of data being used, and potential implications of AI-driven decisions.

Keeping detailed records of the AI system's development process, including data sources, model architectures, and testing procedures, can also contribute to both internal governance and potential external audits.

Accountability

Accountability is essential for ethical AI principles to be implemented successfully. “Computer says no” should not be a response that is provided to any customer or stakeholder. Organisations should clearly define and communicate the roles and responsibilities of different stakeholders in the AI lifecycle so that accountabilities are clear. This includes designating specific individuals or teams responsible for ethical oversight, risk assessment, and compliance. Organisations also benefit from establishing cross-functional ethical review boards to assess high-impact AI projects, so that ethical considerations are integrated throughout the development process.

Employees will not always be naturally aware of how to use AI ethically and responsibly. Therefore, it’s important to conduct organisation-wide training programs to educate employees at all levels about the importance of ethical AI and their role in maintaining it. This creates a shared understanding and commitment to responsible AI practices. This can be reinforced by developing reward structures that incentivise ethical behaviour and decision-making in AI development and deployment. For example, organisations may incorporate ethical considerations into performance evaluations and promotion criteria.



Contestability

Empowering stakeholders means clearly defining and communicating the rights of affected parties to contest AI-driven decisions. This includes understanding and complying with relevant legal and regulatory requirements regarding algorithmic decision-making. Organisations should implement a structured process for human review of contested AI decisions, which should include designated personnel with the authority to investigate and overturn AI-generated outcomes when necessary.

It's also important to provide clear guidance and resources for individuals seeking to contest AI decisions. This may include establishing a dedicated support team, creating user-friendly interfaces for submitting appeals, and offering transparency about the contestation process. Organisations can then use insights gained from contestation cases to refine and improve the AI system, addressing systemic issues that may lead to unfair or inaccurate outcomes.

Reliability and safety

A rigorous data curation process to determine the quality, representativeness, and ethical sourcing of training data can help make AI systems more reliable. This includes addressing potential biases in historical data and guaranteeing diverse representation. Organisations can achieve this by conducting comprehensive pilot studies to assess the AI system's performance, identify potential issues, and refine the model. These studies should involve diverse user groups and scenarios.

Ongoing monitoring mechanisms should be implemented to track the AI system's performance, detect anomalies, and assess its impact on various stakeholders. Regular evaluations should be conducted to identify areas for improvement and alignment with ethical standards.

Organisations should choose or design AI systems with built-in safeguards and fail-safe mechanisms to prevent unintended consequences or system failures. This may include human oversight for critical decisions and the ability to quickly disable or override the system if necessary.

Establishing an AI ethics culture

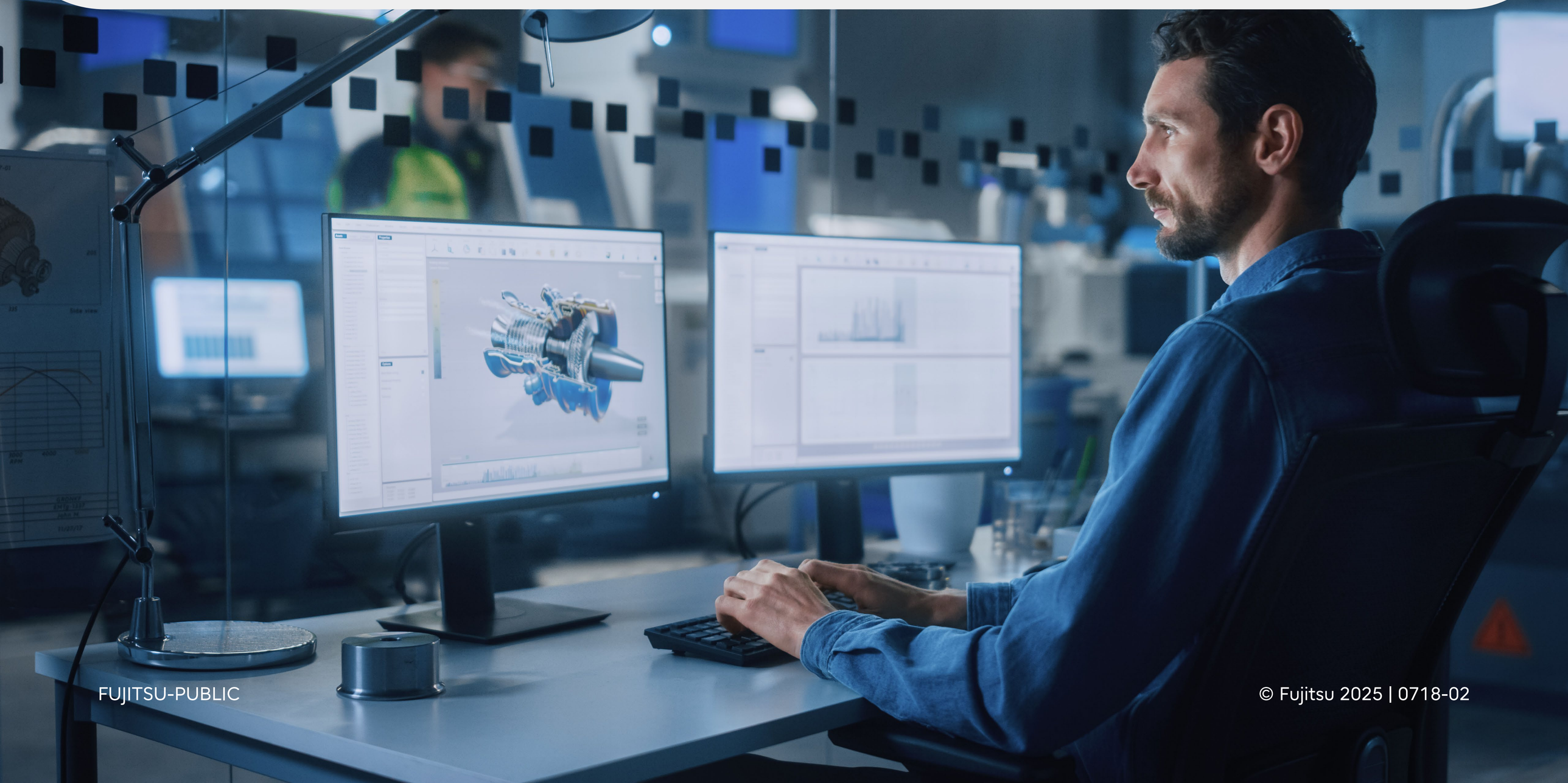
Design

The design stage establishes the foundations of AI ethics at the organisation. This defines the high-level ethical AI principles and policies as well as the governance bodies that will enforce them.

The foundation of an ethical AI culture begins with leadership engagement and commitment. It is crucial to secure buy-in from top executives and board members by educating them on the importance of ethical AI and its potential impact on the organisation's reputation and bottom line. This process should culminate in the development of a clear vision statement for ethical AI that aligns with the organisation's values and mission.

Organisations must also consider appointing a governance board to oversee their AI usage. In some AI-forward organisations, this may be a separate, AI-focused governance board, which is considered best practice. However, in some organisations that are not using AI to a significant extent, a dedicated governance body may not be necessary. In such cases, existing governance bodies can be extended to include AI in their remit.

Organisations must decide what kind of AI culture is appropriate based on what is relevant to their business, markets, customers, and employees. They should also consider the regulatory environment they operate in; for example, a bank may have different AI ethical drivers compared with a marketing firm.



A good starting place is to examine existing frameworks within the organisation. This involves assessing current ethical guidelines, risk management processes, and compliance frameworks to identify gaps and areas where AI-specific considerations need to be integrated.

Simultaneously, a thorough analysis of AI-related laws and regulations in all jurisdictions where the organisation operates should be conducted, potentially engaging legal experts specialising in AI and data protection laws.

Identifying stakeholders is another crucial aspect of the design phase. Organisations should map out all parties impacted by AI systems, including employees, customers, partners, and communities. Conducting stakeholder engagement sessions can provide valuable insights into their concerns and expectations regarding AI use. Part of this should include setting clear guidelines for stakeholder feedback. For example, if an AI solution makes an incorrect, biased, or contestable decision, there must be a clear mechanism for affected entities or individuals to challenge that decision.

Finally, a clear governance structure for ethical AI should be established. This structure should include:

- An AI ethics board or committee with cross-functional representation
- Designated AI ethics officers or champions
- Reporting lines and escalation procedures for ethical concerns
- Integration of ethical considerations into existing risk management and compliance frameworks

Implement

The implementation stage begins with establishing governance groups, including the AI ethics board. These groups should be provided with the necessary resources, authority, and access to oversee ethical AI practices effectively.

Developing compliance processes is the next crucial step. This involves creating step-by-step procedures so that AI solutions comply with ethical policies throughout their lifecycle. These processes may include ethical impact assessments for new AI projects, checklists for different stages of development and deployment, and regular ethics reviews for ongoing AI systems.

Organisations should invest in or develop tools to assist in ethical AI implementation. These tools may include fairness and bias detection software, explainable AI frameworks, and privacy-preserving data analysis tools.

Comprehensive training programs on ethical AI should be developed for various roles within the organisation. These should include basic awareness training for all employees, in-depth technical training for AI developers and data scientists, and specialised training for managers and decision-makers on ethical AI governance.

Once these are in place, it is important to develop a communication strategy to inform all stakeholders about the organisation's commitment to ethical AI. This may include internal communications to employees, transparency reports for customers and the public, and engagement with regulators and policymakers on ethical AI initiatives. Doing so will help cement the AI principles and send a clear message about the organisation’s commitment to ethical AI.

Monitor

Ongoing monitoring is crucial for maintaining an ethical AI culture. Organisations should establish feedback mechanisms, creating channels for employees, customers, and other stakeholders to report ethical concerns or suggest improvements to AI systems. These could include anonymous reporting hotlines, regular surveys on AI impact and perceptions, and community engagement forums.

Regular policy evaluation is essential to ensure the ongoing effectiveness of ethical AI practices. Organisations should review AI ethical policies and their implementation annually or bi-annually, assessing the effectiveness of current policies, identifying emerging ethical challenges, and updating policies and procedures to address new issues.

External collaboration plays a vital role in staying at the forefront of ethical AI practices. Organisations should engage with industry peers, academic institutions, and ethical AI organisations to stay informed about best practices and contribute to the broader dialogue on responsible AI. This collaborative approach makes certain that the organisation's ethical AI practices remain current and aligned with evolving industry standards and societal expectations.

How Fujitsu can help

Establishing an ethical AI culture is not a one-time effort, it is an ongoing journey that requires commitment, resources, and continuous adaptation. It can be complex and challenging, especially as the landscape continues to shift rapidly. Legislators are scrambling to keep up, while business decision-makers are trying to balance the need for speed and efficiency with the requirement to avoid biases and unethical AI use.

Fujitsu has developed an AI Ethics Assessment that supports Australian and New Zealand-based organisations in the journey to responsible and ethical AI usage. It encompasses the following steps:

1. Establish an ethical framework for your business at minimal cost.
2. Ensure fairness, transparency, and accountability in AI decision-making and business practices.
3. Identify and mitigate ethical risks, biases, and their causes using established ethical principles.
4. Build trust for AI solutions to create new business value with a proven method and safeguard your organisational reputation.
5. Meet ethical compliance so that AI systems and applications align with ethical principles, societal values, and legal requirements.
6. Demonstrate a commitment to responsible and ethical AI practices, attracting ethically conscious customers and partners.

FUJITSU

To find out more and to discuss your organisation's AI journey, contact the Fujitsu Data and AI team at: dataandai@fujitsu.com