Fujitsu Australia Limited
Level 7, 664 Collins Street Docklands VIC 3008
Tel +61 (0) 3 9924 3000
Fax +61 (0) 3 9924 3001
ABN 19 001 011 427

27 July 2023

Technology Branch
Department of Industry, Science and Resources
By email: digitaleconomy@industry.gov.au

Submission to *Supporting Responsible AI* discussion paper

Fujitsu welcomes the opportunity to respond to the Federal Government's consultation on *Supporting Responsible AI* and is pleased to present our response.

Fujitsu is a global, purpose-driven company committed to creating a sustainable world and building trust in society. We are also a frontrunner in emergent technologies such as AI, and as part of this place the utmost importance on the responsible design and use of AI. Our perspective is backed by five years of experience embedding AI ethics.

Overview

Balancing regulation and innovation are critical in the context of AI to ensure its responsible, safe and beneficial development. While innovation in AI brings immense potential and transformative capabilities, it also introduces ethical, legal, and societal challenges.

Regulations play a vital role in safeguarding individuals' rights, privacy, and security, as well as promoting fairness, transparency, and accountability in AI systems. This requires collaboration between policy makers, industry stakeholders, researchers, and consumers to shape regulations that promote responsible innovation, ensure transparency, enable public participation, and adapt to the evolving AI landscape. By striking this balance, we can foster an AI ecosystem that maximizes the potential of new technology while upholding fundamental principles of ethics, fairness, and human well-being.

Overall, the questions and answers in the Federal Government's discussion paper encourage discussion around complex nature of AI risk management and offer suggestions for enhancing human-centricity, transparency, accountability, and public trust in AI technologies. The insights presented here serve as a foundation for fostering responsible and ethical AI practices while encouraging innovation and growth in the AI sector.

Fujitsu would welcome further opportunities to contribute to the Federal Government's consultation and policy development processes on AI. For more information on this submission, please contact Fujitsu's Head of Government Relations.

Sincerely

Graeme Beardsell
EVP, CEO Asia Pacific

FUJITSU

Summary of key points

1. The definitions in the discussion paper are generally agreed upon as they are aligned with industry standards.

2. Potential risks from AI not fully covered by Australia's existing regulatory approaches that should be addressed through regulatory development include bias, model drift, privacy and security concerns, accountability risk, and misuse of technology.

3. Non-regulatory initiatives to support responsible AI practices include publishing guidelines and best practices, funding AI ethics research, establishing an ethics hotline, and investing in open datasets for AI research.

4. For effective AI governance across the government, proposed measures include providing AI guardrails, establishing a central ethics review board, and conducting ethical AI training for government employees.

5. Governance measures from other countries are well covered in the document. As an example, Japan incorporates "**Social principles of Human-Centric AI**" as an additional set of principles for consideration.

6. Different approaches should apply to public and private sectors based on goals, data types, accountability, transparency, size, and adoption dynamics. The public sector should set an example in adhering to AI regulations.

7. Additional steps to further support responsible AI practices within the government agencies include whistle-blower protection measures and advocating for internal Ethical AI champions.

8. Both generic solutions as well as technology-specific ones have their specific advantages – with technology specific solutions particularly more suitable for novel and emerging technologies.

9. Transparency throughout the AI lifecycle is critical to mitigate potential AI risks and improve public trust. Mandating transparency requirements across both private and public sectors would be essential to achieve these goals.

10. Regulating based on risks, rather than banning is the suggested approach.

11. To encourage public trust in AI, public engagement, regulation, transparency, and accountability are essential.

12. Regarding the impact of banning high-risk activities, an assessment based on risk, benefit, and any negative impacts based on lack of innovation would need to be made.

13. Australian conformity infrastructure has existing laws to mitigate some AI risks, but more needs to be done. Existing laws may need to be reviewed and strengthened, and new legislation may need to be developed.

14. Fujitsu supports a risk-based approach to addressing potential AI risks.

15. A risk-based approach is seen as a key step in preventing harm within society.

16. Although a risk-based approach should be applied across all organisations, it might benefit certain organisations more based on the industry sector, size and their AI maturity.

17. Fujitsu supports the elements of risk-based approach presented in Attachment C.

18. Modifying existing frameworks to include AI risk-based approaches is the preferred option, with the suggestion of using AI itself to improve these frameworks.

19. Risk management frameworks should apply to all types of AI systems including large language models (LLMs) and multimodal foundation models (MFMs).

20. A hybrid approach that uses a combination of both voluntary and industry-led tools might be the most effective solution for regulating AI.

Section three provides more detailed answers to the twenty questions provided in the discussion paper.

Fujitsu's approach to responsible AI

Fujitsu is a Japanese multinational information and communications technology (ICT) company headquartered in Tokyo, Japan. It is one of the top 10 largest IT services providers in the world by annual revenue, and the largest in Japan. Fujitsu has approximately 130,000 employees supporting customers in 180 different countries. Fujitsu Australia was established over 50 years ago and is one of the leading ICT providers in the country with a revenue of over 1 billion dollars.

Fujitsu designs and offers broad range of IT and digital products, services and solutions, we invest heavily in R&D and commercialisation of emerging technologies such as quantum inspired computing and AI and provide many essential services to both public and private sector clients in a variety of industries.

We operate our business through the lens of our purpose, which is to "make the world more sustainable by building trust in society though innovation". Our aim supports a collective approach to sustainable transformation. We strongly believe that technology, and more specially AI, will be critical to the acceleration of solutions to the world's most complex societal and environmental challenges.

We also recognise that AI is not free from risk, as we address in our response. As a result, Fujitsu has been an early adopter of AI ethics principles and participation. Fujitsu has advocated for a human-centric approach to technology since 2009 and is a founding member of AI4People, established in 2018, which created an ethical framework for good AI in society.

Fujitsu first launched the **Fujitsu Group AI Commitment** in March 2019, which is a set of 5 principles, based on the bioethical principles, that outlines our commitment on AI:

- Provide value to customers and society with AI.
- Strive for Human Centric AI.
- Strive for a sustainable society with AI.
- Strive for AI that respects and supports people's decision-making.
- Corporate responsibility, emphasise transparency and accountability for AI.

In addition to establishing principles, a key priority for Fujitsu has been governance. In support of this, we have established:

- the **Fujitsu Group AI Ethics External Advisory Committee** in 2019, a committee of prominent external experts that we consult with regularly for the purpose of objectivity. Both our CEO and COO also sit on this committee. We also actively participate in discussions on AI Ethics around the world, including the Global Partnership on AI and standardizing bodies such as ISO.

- In February 2022, **Fujitsu established the AI Ethics and Governance Office (AIEGO),** tasked with actively promoting and governing AI Ethics within the organisation. In Japan, the company also provides an AI ethics e-learning program for all employees, which will be rolled out worldwide.

To compliment our AI governance, Fujitsu uses tools and techniques to help its employees, partners and customers address some of the issues associated with AI Ethics. Some tools of importance to note are:

- Fujitu's AI Ethics Impact Assessment: Our AI Trust Research Center developed "AI Ethics Impact Assessment", a resource toolkit offering developers guidance for evaluating the ethical impact and risks of AI systems based on international AI Ethics guidelines. We offer these resources free of charge to promote the safe and secure deployment of AI systems in society.

- Fujitsu's Fairness Checker: This allows us to detect and mitigate "intersectional bias" caused by various combinations of attributes. The Fairness Aware Machine Learning (FAML) tool allows users to compare different bias mitigation techniques visually and choose the best one.

Responses to discussion paper questions

Although there are a number of laws and regulations that may apply to the use of AI in Australia, including the Privacy Act 1988, the Australian Consumer Law, and the Competition and Consumer Act 2010, AI introduces new risks and challenges beyond the scope of existing regulations. Some of the potential risks resulting from the use of AI include:

- **Bias:** DTA guidance talks about "*deliver impartial and just treatment or behaviour without favouritism*", while the AI Ethics Principles talks about "*Fairness*" – all of this alludes to **bias**, and this is a big risk that needs to be addressed and perhaps there is more to be done to ensure that data and algorithms are designed to preclude bias based on protected characteristics outlined in anti-discrimination laws.

- **Model Drift:** Model drift starts to occur when the performance of the AI system starts to degrade over time, as the data and/or the behaviour of the model starts to change. There is a constant need to retrain the model so that this doesn't occur or lead to unintended consequences. Model drift requires frequent monitoring, retraining and human intervention/validation so that the system behaves as expected. This is especially important when AI is combined with Automated Decision Making (ADM).

- **Privacy:** Due to the ingression of vast amounts of training data from a varied sources, including the internet and social media, AI amplifies the already present risks regarding privacy of individuals in the digital age. The Privacy Act review by the Attorney-General's Department is a welcome step towards addressing some of these issues, however regular reviews are encouraged in the light of rapidly emerging technologies.

- **Security:** AI introduces new vulnerabilities which current security regulations are not equipped to handle. Systems are exposed to new attack vectors enabled by Generative AI such as Adversarial Attacks where AI systems are corrupted by feeding them malicious data or exploiting vulnerabilities in the underlying algorithms. Similarly, data poisoning attacks manipulate the training process and introduce vulnerabilities or biases into the AI system.

- **Accountability Risk:** Who is held accountable in the regulatory framework?

- **Misuse of technology:** More can be done to prevent misuse of the technology, for example by generating false news and deep fakes, identity theft, misuse of facial recognition, interference with the democratic system etc. are risks that lead to systemic societal issues and significantly degrade trust in AI.

Further non-regulatory steps that may be able to assist

Further non-regulatory initiatives the Federal Government could implement to support responsible AI practices in the country:

- **Publishing guidelines, best patterns, and practices**: Provide tools (such as the "AI Verify" tool from Singapore) and checklists that businesses can download and use. Additionally, create accreditation and certification programs for businesses.

- **Funding research in AI Ethics:** To help in bias mitigation, analysing drift, explainable AI, etc., that in turn contributes to the guidelines.

- **Ethics Hotline:** Establish an ethics hotline where citizens can report any concerns or issues related to the use of AI. This hotline can serve as a platform for citizens to voice their concerns and seek guidance on ethical implications and potential risks.

- **Publishing open datasets:** The Federal Government could also consider investing in open datasets which are 1) representative of the population, 2) free of bias and 2) respectful of privacy and data protection. This would support AI research and development of responsible AI practices in a way which is more trustworthy and transparent.

- **Programs to encourage diversity within the AI industry**: Generative AI technology has a high barrier to entry, due to considerable educational and training requirements, it is likely going to experience under representation from diverse demographic groups – making it increasingly hard to remove bias and achieve fairness. Representation of the community serviced will encourage inclusive and fair AI design and application and broadens the talent pool and increases business's capacity.

- **Programs in place to ensure that accessibility is fairly distributed**: AI tools require considerable internet bandwidth, power and suitable devices, which are not available or affordable to everyone. Regional Australians and older Australians particularly experience poorer digital inclusion and will suffer accelerated digital divide due to this emerging technology. Investment in programs which close the digital gap will ensure that no Australian is left behind.

- **Establish a mechanism for critical reflection:** We encourage critical reflection earlier in the process. Introspective can be less effective, despite best efforts, those developing algorithms will be prone to bias and intellectual lock-in. People should not judge their own work, or the work where there may be a conflict of interest. Critical audiences that opposes algorithms and points out their shortcomings will be essential in the development of quality, responsible AI.

Investing in these initiatives could facilitate the following benefits:

- Boost the economy by creating additional jobs and new business opportunities.
- Help build trust in AI and ensure that it is used in a reasonable and fair manner within legal frameworks.
- Help to prevent misuse of AI.

Suggested coordination of AI governance across government

- **Provide AI guardrails:** Establish a set of guardrails, which are a set of rules and checklists that need to be used while evaluating and using AI. These guardrails should also include a set of do's and don'ts, best practices and guidelines for the use of AI in any government agency.

- **Establishment of a Central Ethics Review Board:** Getting different government agencies to create their own review board could lead to inconsistencies and confusion. A more streamlined and cohesive approach would be to establish a central ethics review board that is responsible for evaluating and approving the use of AI technologies across all government agencies. This would assist in assessing the ethical implications, risks, and potential societal impact of AI deployments, ensuring compliance with ethical guidelines.

- **Ethical AI Training:** Provide comprehensive training programs to government employees on ethical AI practices, including topics such as bias mitigation, fairness, privacy, and accountability. This training would equip employees with the necessary knowledge and skills to ensure responsible use of AI in their work. This will also help in alleviating any fear over the use of AI.

Some of the priority aims of a central approach could include:

- **Responsible use:** Ensure that AI is used in a responsible and ethical manner and would benefit society.

- **Minimise risk**: Minimise the risk of AI introducing bias as well as misuse.

- **Leader in AI**: Ensure that Australia doesn't get left behind in the uptake of AI and that we remain world leaders in this space.

International governance examples

The discussion paper acknowledges other examples across the EU, USA, UK, Canada, China, NZ, Singapore Thailand, Italy and Indonesia. Some additional examples to note include:

- **AI regulatory landscape in Japan:** Japan currently takes a voluntary, risk-based, AI-agnostic approach to regulation of AI. There it is the sentiment that regulation can struggle to keep up with the rapidly changing AI landscape, and therefore are opting for more of an "agile governance" based approach will be more effective and responsive to change. They also advocate for multi-stakeholder collaboration. Stakeholders include not only experts in technology, law, economics, and management but also individuals and communities as the ultimate beneficiaries of AI governance.

- Japan, much like Australia, have laws that are relevant to the development and use of AI which address privacy, risk disclosure etc. The Japanese Government provide a number of guidelines for implementation of ethics principles and topics related to responsible AI. They also advocate for multi-stakeholder collaboration. Stakeholders include not only experts in technology, law, economics, and management but also individuals and communities as the ultimate beneficiaries of AI governance.

- **Japan's Social Principles of Human-Centric AI:** The cabinet office in Japan has issued "**Social principles of Human-Centric AI**". This document provides a set of principles that are based on three basic philosophies: human dignity, diversity and inclusion, and sustainability. The goal of the Social Principles is not to restrict the use of AI in order to protect the principles, but rather to realize them through AI. This corresponds to the structure of the Organization for Economic Cooperation and Development's (OECD) AI Principles, with the first principle being to achieve "inclusive growth, sustainable development, and well-being" through AI.

To achieve these goals, the Social Principles set out seven principles surrounding AI: human-centric; education/literacy; privacy protection; ensuring security; fair competition; fairness, accountability, and transparency and innovation. It should be noted that the principles include not only the protective elements but also the principles that guide the active use of AI, such as education, fair competition, and innovation. This is also reflective of Japan's regulatory reform to allow for adoption of AI on a case-by-case basis, for example, revision of the Road Traffic Act and Road Transport Vehicle Act to allow for Level 3 automated driving and establishment of the "Digital Rincho"; a taskforce with the aim of revising regulations.

- **Support from Institutions:** There have been a few recently established institutes that aim to position themselves an an authority for governments draw expertise and support from. Examples include Montreal AI Ethics Institute in Canada, Global AI Ethics Insight in France, The Oxford Institute for AI Ethics in the UK, and Gradient Institute in Australia. Some of these bodies are already partnering with government entities.

- **Adopting global standards:** It is encouraging to see Australia as a signatory to the OECD's AI principles, and its engagement in bilateral, regional and multilateral discussions with other jurisdictions, where significant work on AI is being undertaken. This includes the OECD, United Nations, World Trade Organisation and the World Economic Forum (WEF).

- **Use of sandboxes:** The discussion paper briefly discusses how other countries are planning to enable regulatory levers (e.g., AI sandboxes). Sandboxes could contribute to the Australian policy environment and would support the transition from research to development of appropriately developed AI systems.

How can the Australian Government further support responsible AI practices in its own agencies?

In addition to what has been highlight in Q4, we offer the additional feedback:

- **Whistle-blower Protection:** Implement robust whistle-blower protection mechanisms to encourage government employees to report any unethical or irresponsible use of AI within their agencies. This protection would ensure that employees feel safe and supported when exposing potential AI-related misconduct.

- **Internal Ethical AI Champions:** Identify and empower internal champions within government agencies who are passionate about ethical AI. These champions can serve as advocates and influencers, promoting ethical AI practices and ensuring their adoption within their respective agencies.

- **Foster collaboration and knowledge sharing amongst agencies**: For example, the UK Government's Office for Artificial Intelligence and the UK AI Council act as central coordinating bodies, encouraging collaboration and knowledge sharing among government departments.

- **Embed legal and ethical assessments into decision making**: embedding impact assessments into usual decision making, or taking an Ethics by Design approach to AI, will help to ensure that rigorous governance is upheld by its own agencies.

In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better?

- Generic solutions can be advantageous when it comes to critical applications in the areas of healthcare, finance and transport. Generally, technology-specific solutions are advantageous when it comes to novel applications.

- While generic solutions are valuable to ensure consistency and avoid conflicts across sectorial regulations, technology-specific solutions can be employed for emerging technologies (e.g., autonomous motor vehicles) to facilitate innovation (providing exemptions from some existing regulations, for instance).

Input on where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI, and mandating transparency requirements across the private and public sectors.

Transparency is important in all stages of the AI lifecycle, from data collection to model deployment. The reasons for this include:

- **Building trust:** When users know how their data is being used, and decisions are made by the AI, it improves trust and contributes to the success of the project.

- **Mitigating bias:** Transparency can help identify bias and mitigate against it.

- **Accountability:** When users know how the AI system works, they are more likely to hold companies accountable. This will in turn ensure that they are used in a responsible manner.

- **Explainability:** Explainability of AI systems increases the trustworthiness of these technologies. AI can be instrumental in communicating reasoning to consumers as to why things are serviced the way they are. And for consumers who are not realising that there's an AI capability behind their services, it's critical to be able to explain why suggestions are made in certain ways or what logic is there behind their services.

What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

Public trust in AI can be increased by:

- **Explainability** – When AI systems can clearly explain how decisions were made, it will not only help consumers understand the system, but also help in mitigating bias.

- **Public engagement** – Initiatives to increase people's awareness of AI through education, and public awareness campaigns to remove any fear associated with the use of AI.

- **Regulation** – When consumers are aware that checks and balances are present in the form of legislation, their trust in AI systems would be better supported. In addition, there must be clear accountability and compliance enforcement.

- **Transparency** – In addition to transparency of decisions (through explainability), ensuring that consumers know when an AI is being used is also critical.

How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?

Banning some AI technologies may incur the following impacts:

- **Balance of risk aversion vs openness to innovation** – If the risk aversion limits Australia's innovation capacity, this may stimy Australia's competitiveness in a number of trade sectors.

- **Risk vs benefit compromises** – If there is a high positive impact linked to high-risk scenarios, there can be some evaluations as to how this risk can be addressed in different ways instead of banning the technology completely. Some automated systems that analyse the risks and benefits of these technologies could also assist with the assessments.

What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?

Currently, there are a number of laws and standards within the Australian conformity infrastructure that could potentially impact AI usage or mitigate against AI risks.

These laws could be used to regulate potential risks with AI systems such as bias, misinformation, and improper use of personal information. However, more needs to be done in this space. Here are some suggestions to improve the Australian conformity infrastructure to mitigate potential AI risks:

- **Review and strengthening of existing laws and regulations:** Some of the existing laws may not address more recent risks that have emerged from the use of AI. Reviewing and strengthening them would go a long way towards ensuring that the risks are mitigated.

- **Development of new legislation:** Existing laws may not go far enough to address some of the risks posed by AI. As a result, new standards would need to be developed to specifically address issues such as fairness, bias, accountability, security, privacy, etc, as well as penalties for non-compliance.

Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

A summary of our view on this subject is as follows:

- **Risk Management** – We interpret “risk-based” as risk management in a wider scope. Under this interpretation, risk-based approach is acceptable or even preferrable because this approach allows us to consider risks with different characteristics and benefits to assist in minimizing adverse consequences of those risks. In the context of EU AI Act, “risk-based approach” refers to the categorisation of applications to one of unacceptable, high-risk, low-risk, and so on.

- **Key Risk Indicators** – KRIs or *Key Risk Indicators* are a way to proactively measure risk and this could also be applied in the context of risks associated with AI. KRIs can be anything from the number of customer complaints related to tasks being completed by AI to the number of software errors. By monitoring KRIs, early warning signs of potential risks can be more readily apparent. The development, monitoring, and analysis of KRIs is a data-driven and intelligent approach to risk management.

- **Risk Managing for AI and AI for Risk Management** – Risk Management for AI is AI for Risk Management. But these two can be brought together. While managing all AI risks present similar principles to any other risks, AI can be used to better manage its own risks. AI can not only automate all deviations from risk boundaries, it can also help analyse causational and correlational relationships with risks, and its decomposed factors. AI can help monitor KRIs to get even earlier warning signs of risks and then determine the most impactful ways to address them.

What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

Under the same interpretation as above, risk-based approach could be the only option to prevent harm from occurring as a result of technology and the use of AI. Some of the main benefits of risk based approach are:

- **Proactive Risk prioritisation :** This approach enables us to prioritise risks even before deployment of AI solutions. Because of this, it allows us to take measures to mitigate, monitor and review risks proactively.

- **Consistency:** A risk based approach creates a consistent and structure-based approach to evaluating AI solutions.

- **Turning risk into opportunity:** Mitigating AI risks can potentially lead to new innovation opportunities, as a low priority item could suddenly be elevated to a higher priority item as remedy to high risk situation.

- A key limitation of a risk based approach is that risks are often challenging to quantify precisely as expert evaluations and decisions can be influenced by uncertainties. Additionally, strictly adhering to risk considerations may potentially decrease customer value.

- Nonetheless, we believe that addressing this concern can be achieved through design approaches and integrating risk management seamlessly into the development process.

Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

A risk based approach should be applied equally to all organisations, however there are a few considerations that should be taken into account:

- **Sector -** Some sectors by nature have a higher effect on consumers such as health and finance. While going through risk/ benefit assessments, this needs to be taken into account. Suggesting a wrong medical treatment to a patient has a considerably higher risk than suggesting a wrong clothing option, for instance. Where risks are potentially severe, a robust and intelligent risk-based approach will be more beneficial.

- **Maturity and resources** – By following a risk based approach, organisations can ensure that they are only taking on the tasks that are necessary to address the risks associated with AI systems. This can help to reduce the burden on organisations and ensure that they are only focusing on the most critical risks. This could be particularly beneficial to organisations who do not have any significant AI capability or do not have enough maturity with their AI adoption.

What elements should be in a risk based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

We support the elements presented in Attachment C. Our view is that it is preferrable to evaluate both likelihood and impact of risks in a risk-based approach.

How can AI risk based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

- **Incorporating into existing frameworks** - By adding an AI risk based approach to existing assessment frameworks and risk management processes, organisations can make their work more efficient and avoid duplication. Modifying the existing assessment frameworks to include AI-specific risk assessment criteria could involve adding new sections or augmenting existing ones with considerations relevant to AI systems.

- **Frameworks for AI and AI for frameworks** – Another approach that may be useful would be to consider how AI can be used to improve existing frameworks. For example, AI can be used to identify misalignments with privacy policies, or to identify factors that impact the consistent and persistent management of privacy. This can be extended to other uses of AI, such as identifying Key Risk Indicators (KRIs) and linking them with existing assessment frameworks. This will make assessments easier, more intelligent, and more automated.

How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?

- **Risk management frameworks are technology agnostic:** This means that they are not specific to any particular type of technology. They can be used to assess and manage risks associated with a wide range of technologies, including AI systems such as LLMs and MFMs.

- **The impact of risks posed by LLMs and MFMs can be significant**: This is because these systems are capable of generating large amounts of text, code, and other content. This content could be used to create fake news, spam, or other harmful content. It could also be used to generate text that is biased or discriminatory.

- **However, the nature of risks posed by LLMs and MFMs is not fundamentally different from the risks posed by other AI systems:** This is because the risks associated with these systems - such as bias, fairness, safety, security, and privacy, are ultimately derived from the same underlying risks associated with AI systems in general.

In summary, risk management frameworks are technology agnostic, so they apply to all types of AI systems, including LLMs and MFMs. The impact of risks posed by LLMs and MFMs can be significant, but the nature of these risks is not fundamentally different from the risks posed by other AI systems.

While following the same risk management framework, additional steps can be taken specific to LLMs and MFMs in the near future:

- **Partnering with leading institutes:** There may be some challenges due to the novelty of these technologies, which may raise questions about the assessments' validity. However, partnering with ethical and research-focused institutes can help to address this issue.

- **Close monitoring of LLMs and MFMs** – Given the relatively new nature of LLMs and MFMs, it is prudent to closely monitor and supervise their applications to understand the associated risks better. Regular reports can be mandated as part of this supervision. Initially, focusing on straightforward risks can aid in monitoring, while this process can also uncover additional, less obvious risks. As the technologies mature and become more familiar, this monitoring can help build confidence in their responsible use.

Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to: a) public or private organisations or both; b) developers or deployers or both?

Both voluntary/ self-regulation and mandated regulation have their advantages and drawbacks:

- **Voluntary and industry-led regulation** allows organisations to have more flexibility and control over how they implement risk-based approaches, while being more cost effective. It can also foster innovation and flexibility as organisations can implement tailored strategies. However, this is difficult to enforce, and may be difficult hold organisations accountable for non-compliance. It may also lack consistency across the industry.

- **Mandatory regulation** on the other hand provides a clear and uniform set of standards for all organisations to comply with. The main drawback is the higher cost involved and the approach can be inflexible and less responsive to the needs of evolving developments and technology.

The risk-based approach for responsible AI should be applied to all organisations, whether they are public or private; whether they are the developer or deployer. This is because, irrespective of the role they play, they have an equal responsibility to ensure that AI systems are used in a responsible and ethical manner.

Thank you for the opportunity to present our submission on this important discussion paper.

References and links

- AI Ethics and Governance : Fujitsu Global
- Japan's Social Principles of Human-centric AI:
- Montreal AI Ethics Institute
- Oxford Institute of Ethics in AI
- Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency (csis.org)
- Ministry of Internal Affairs and Communications
    - Information and Communications Policy Research Institute|About Us
    - The Draft AI R&D GUIDELINES for International Discussions
    - The Conference toward AI Network Society