



24 July 2023

Department of Industry & Science  
10 Binara Street,  
Canberra ACT 2600

**Re: Safe and Responsible AI in Australia discussion paper**

Salesforce welcomes the Australian Government's consultation on Safe and Responsible AI. It is a considered approach that takes into account international developments and Australia's unique opportunities and challenges.

**About Salesforce**

Founded in 1999, Salesforce is a global leader in cloud enterprise software for customer relationship management ("CRM"); providing software-as-a-service ("SaaS") and platform-as-a-service ("PaaS") offerings to businesses, governments and other organisations around the world. Our customers represent companies of all sizes and across all sectors. Our business model is cloud-based and low code, allowing for faster deployment of technologies and greater agility. We help our customers connect with their customers — or employees or citizens — in a whole new way using cloud, data and AI technologies.

**Salesforce & AI**

Salesforce's AI capability, called "Einstein," is built into the Salesforce platform, and is designed to combine artificial intelligence with Salesforce's suite of enterprise services, democratising the power of AI for every Salesforce user. Many of the use cases for both our AI and generative AI are for creating efficiencies at the enterprise level. Some current examples consist of suggesting a next best step or drafting an introductory email for a sales lead. These use cases are generally lower-risk and are augmenting human decision making processes.

We believe that the tremendous benefits of AI should be accessible to everyone, while ensuring that those technologies remain safe and inclusive. At Salesforce, we are committed to providing our employees, customers, and partners with the tools they need to develop and use AI safely, accurately, and ethically. Our commitment to [ethical AI](#) consists of adherence to the following principles:

- **Responsible:** To safeguard human rights and protect the data we are entrusted with, we work with human rights experts, and educate, empower and share our research with customers and



partners to enable them to use AI responsibly. We strive to comply with the laws and values of the markets in which we operate. We strive to adhere to the highest security and safety protocols.

- **Accountable:** Accountability to customers, partners and society is essential. Independent feedback should be sought for continuous improvement of practices and policies and work to mitigate against harm to customers and consumers. We seek stakeholders' feedback, take guidance from our [Ethical Use Advisory Council](#), and conduct our own data science review board.
- **Transparent:** Our customers should be able to understand the “why” behind each AI-driven recommendation, output and prediction so they can make informed decisions, identify unintended outcomes, and mitigate harm. We [strive](#) for model explainability and clear usage terms, and ensure customers control their own data.
- **Empowering:** AI is best utilised when paired with human ability, effectively augmenting people and enabling them to make better decisions. Accessible AI promotes growth and efficiency, and benefits society as a whole.
- **Inclusive:** AI should respect the values of all those impacted, not just of its creators. To achieve this, we test models with diverse data sets, seek to understand their impact, and build inclusive teams.

Because of the rapid evolution of the technology, along with the opportunities and challenges emerging from the use of generative AI, we have gone one step further and articulated an additional set of [guidelines](#) meant to guide the development of trusted and responsible generative AI, at Salesforce and beyond.

- **Accuracy:** We need to deliver verifiable results that balance accuracy, precision, and recall in the models by enabling customers to train models on their own data. We should communicate when there is uncertainty about the veracity of the AI's response and enable users to validate these responses.
- **Safety:** As with all of our AI models, we should make every effort to mitigate bias, toxicity, and harmful output by conducting bias, explainability, and robustness assessments, and red teaming. We must also protect the privacy of any personal data present in the data used for training and create guardrails to prevent additional harm.
- **Honesty:** When collecting data to train and evaluate our models, we need to respect data provenance and ensure that we have consent to use data. We must also be transparent that an AI has created content when it is autonomously delivered.
- **Empowerment:** There are some cases where it is best to fully automate processes but there are other cases where AI should play a supporting role to the human — or where human judgment is required. We need to identify the appropriate balance to “supercharge” human capabilities and make these solutions accessible to all (e.g., generate ALT text to accompany images).

- Sustainability: As we strive to create more accurate models, we should develop right-sized models where possible to reduce our carbon footprint. When it comes to AI models, larger doesn't always mean better: In some instances, smaller, better-trained models outperform larger, more sparsely trained models.

As AI becomes ubiquitous in the economy, policy makers and industry should work together to establish guardrails ensuring the ethical development and utilisation of this powerful tool.

Salesforce is committed to building trusted, transparent, and accountable AI systems that prioritise fairness, accuracy, privacy, and positive societal impact.

Salesforce supports risk-based AI regulation that differentiates contexts and uses of the technology (e.g. B2C v. B2B) and assigns responsibilities based on the different roles that various entities play in the AI ecosystem. AI regulation is most effective when built upon, and interoperable with, robust privacy standards. One-size fits all approaches that aim to regulate the technology in a uniform way rather than mitigating the risks from specific uses of the technology, would hinder innovation and competition, and delay the adoption of technology that is already being productively used by consumers and businesses around the world.

Below are Salesforce comments to the discussion paper:

### **Definitions**

We welcome a common approach to definitions and obligations that will create more durable, robust, and interoperable norms.

Salesforce supports efforts to create common standards such as Standards Australia SA TR ISO/IEC 24027:2022, Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making. This standard provides requirements to help ensure AI technologies and systems meet critical objectives for functionality, interoperability and trustworthiness. SA TR ISO/IEC 24027:2022 is an identical adoption of the ISO/IEC TR 24027:2021 standard, and also specifies the measurement techniques and methods for assessing bias with the intention to address bias-related vulnerabilities.

### **Responses suitable for Australia**

We note Australia's Responsible AI Network will act as a gateway for Australian industries to uplift their responsible AI practices.

Salesforce has previously advocated the Australian government consider establishing an expert Advisory Council to help in providing strategic advice in relation to AI, examples include the NSW Artificial

Intelligence (AI) Advisory Committee which comprises of experts to use their extensive and varied experience to provide ongoing strategic advice on the use of AI to assist in decision-making and improving service delivery across the NSW Government.

Another example is Singapore's Advisory Council on the Ethical Use of AI and Data which was formed in 2018 for the purpose of advising Singapore Government on ethical, policy and governance issues arising from the use of data-driven technologies in the private sector and providing general guidance to businesses to minimise risks, and to mitigate the adverse impact on consumers from harm.

Salesforce supports governments and industry working together, like the [work occurring](#) in the G7, G20, the Organisation for Economic Co-operation and Development (OECD), or the U.S. National Institute of Standards and Technology (NIST).

### **Target areas**

Although facial recognition technology can have important uses subject to meeting privacy and security requirements, such as when individuals voluntarily use it to unlock devices (e.g. our phones), it can also pose challenges. Salesforce has taken the position of not allowing the use of our products for facial recognition.

We have done so by considering how we develop our AI technologies and the safeguards we develop for their responsible use. Concerns remain about facial recognition technology, both around its accuracy and the harm it can cause, particularly to minority communities. Error rates can climb sharply when factors like aging, camera viewpoint, distance, barriers, illumination, shadows, and movement are introduced.

Other target areas for consideration:

- Privacy issues and data mining: Companies need to have transparency around where they're gathering data and how they're using it.
- Copyright concerns: Given generative AI tools pull from vast data sources, the chance of plagiarism increases.
- Misinformation: False information could [spread more quickly](#) with AI chatbots, which also have created entirely inaccurate stories called [hallucinations](#).
- Identity verification: There is the [need to verify](#) articles, social media posts, art, and more.
- Child protection: There's been a call to ensure children and teenagers are protected against [alarming, AI-generated content](#) on social media.

### **Potential gaps in approaches**

As Salesforce has advocated in previous submissions, the Australian Government should consider appointing a "Chief ethical and humane use officer" or AI Safety Officer, to help with the implementation of the framework for the ethical and humane use of technology within the Australian government and as it applies to digital government services.

Unlike the eSafety Commissioner which has a regulatory function across industry, the proposal for an AI Safety Officer is to ensure the ethical use of AI within government.

Other models for Australia to consider is AI Verify, a subsidiary of the Infocommunications Media Development Authority of Singapore (IMDA). It is an AI governance testing framework and software toolkit that validates the performance of AI systems against a set of internationally recognised principles through standardised tests, and is consistent with international AI governance frameworks. Salesforce is one of the founding members.

### **Risk based approaches**

Salesforce supports tailored risk-based AI regulation that differentiates contexts and uses of the technology and ensures the protection of individuals, building of trust, and innovation.

A tailored approach is key: A one-size-fits-all approach to regulation may hinder innovation, disrupt healthy competition, and delay the adoption of the technology that consumers and businesses around the world are already using to [boost productivity](#). For example, whereas Salesforce encourages responsible AI development for all use cases, a small team of engineers developing a generative AI chatbot to help people with completing an application form should have fewer guardrails and oversight than healthcare providers using AI tools used to help diagnose patients and develop care plans.

We commend policymakers who apply a nuanced approach when developing regulation, and to consider the following concepts to help society navigate this important moment:

- Risk-based framework: The context in which technology is used matters, and some industries, like healthcare, are more likely to create higher risks for their users and society than others. [Risk-based AI regulation](#) would focus most on high-risk applications, especially those that could cause significant harm or impact someone's rights and freedoms.
- Differentiation based on context, control, and use: Regulation should differentiate the context, control, and uses of the technology and assign guardrails accordingly. Generative AI developers, for instance, should be accountable for how the models are trained and the data they are trained on, while those deploying the technology decide how the tool is being used and should have rules according to that interaction.
- Data privacy laws: [Data protection laws](#) that protect the fundamental human right to privacy are a foundation of responsible AI regulation. AI is powered by data. Additional rules specific to

generative AI should address the use and privacy of personal data for training future models, safeguarding personal data within the AI ecosystem.

- Transparency: AI systems might operate as "black boxes," making it difficult to understand their decision-making processes. Individuals should be informed of and understand the "why" behind AI-driven recommendations, and they should be aware if they are interacting with a human or a simulated persona.
- Accountability and government oversight: AI impact assessments are one way to promote accountability and trust with high-risk AI systems. Licenses or notifications can serve a useful role with accountability and compliance, but should be implemented through a risk-based approach and balance compliance with market entry, competition, and innovation. That's why risk management frameworks like those shared by [NIST](#) will bring a foundational understanding to the field.
- Harmonisation and consistency with existing rules: Many existing laws and policies already provide some guardrails around AI, such as [global data protection laws](#). As regulators and other stakeholders develop new guidance, they should assess and clarify whether there is an existing law addressing these concerns.
- Future-proof and universal applications: Given the pace of innovation, AI regulations be globally interoperable, and be both durable and flexible, focused on the ethical development and deployment of AI systems rather than focused on a specific technology at a specific time.

Thank you for the opportunity to comment on the discussion paper.

Yours sincerely

A handwritten signature in black ink that reads "Sassoon Grigorian".

Sassoon Grigorian

**VP, Government Affairs & Public Policy, APAC**