



## RECOMMENDATIONS

### Preferred definitions

1. Artificial Intelligence (AI) is a field of science concerned with building computers and machines that can reason, learn, and act in such a way that would normally require human intelligence or that involves data whose scale exceeds what humans can analyse.
2. Machine Learning (ML) is a subfield of AI, a program or system that trains a model and learns patterns from input data to make useful predictions from new or never before seen data. ML gives the computer the ability to learn without explicit programming.
3. Large Language Model (LLM) is a type of generative AI that specialises in the generation of novel combinations of text in the form of natural sounding language.

### AI governance

4. The Commonwealth and States should agree on a single consistent approach to AI regulation, working together on common objectives, standards and practices.
5. If AI governance is proposed for the private sector, it should also apply to the Commonwealth and state governments.

### Transparency and trust

6. Building trust is dependent on lifting AI literacy, educating people on AI benefits and risks, and putting in place clear processes for consumers wanting further disclosures about the AI involved in a transaction.
7. As provided by Equifax today, consumers are given an explainability of their credit score band that is specific to the individual's behaviours. This makes it both understandable and actionable.
8. Beyond the credit reporting regime, the level of transparency desirable needs to be considered in the context of the given application, the impacted stakeholder and the potential for unintended consequences.

### Risk based approach

9. With data already subject to multiple different legal frameworks, there is a risk another piece of legislation will introduce overlap, confusion and/or contradiction. Any new approaches should be designed and implemented to avoid duplication.
10. Because the field is evolving so rapidly, legislation is unlikely to keep pace, hence a supervisory or regulatory agency should be considered to continuously monitor and review developments in a way that has not been necessary previously.

11. There is ambiguity in how the proposed risk levels are currently defined. Work is needed to elaborate the use cases and take into account the likelihood of harm in addition to the severity of the harm. A suggestion to complement the elements presented in attachment C could be “AI in the loop assessments” prior to ‘human in the loop’. This can be considered as ‘AI to mitigate AI risks’ or ‘AI to audit AI’ to address ‘scalability’, which is one of the main challenges of ‘human in the loop’ assessments.
12. Additionally, the current definitions for the requirements of explanation for medium and high risk are problematic for both organisation and consumer. Detailed AI outputs create intellectual property risks and may bring a level of detail and technicality that is not actionable or informative to consumers.

## **Equifax**

Equifax is a data, analytics and technology company, best known as Australia's leading provider of consumer credit reporting information. We use data, industry expertise and innovative analytics, including AI, to transform knowledge into insights that help lift confidence in decision making.

### **Credit reporting and AI**

What can be viewed as one of the first examples of regulating algorithmic output occurred in 2021, when a new law required credit reporting bodies to disclose to consumers their credit score band and provide them with the key contributing factors characterising that score band.

Equifax supported this, believing that "one of the greatest limitations of AI systems is a lack of visibility around how decisions are made. In contrast, explainable AI (xAI) seeks to produce transparent explanations of how conclusions are reached."

Transparency boosts confidence in the AI used in credit scoring and to this end, Equifax exceeds the generic requirements of the law (what are the key contributing factors typical for people to be in a score band) by providing consumers with four key contributing factors, in detail, tailored to the person and setting out steps the individual can take to improve their credit risk standing.

We are able to do this because of developments in AI techniques. Equifax's NeuroDecision® Technology is the first patented credit scoring system to use explainable AI. It uses a neural network algorithm to predict the probability that a consumer will default on a credit product.

Because NeuroDecision® Technology is fully interpretable, lenders can realise the benefits of AI without introducing a 'black box' algorithm that could reduce their control of credit risk. Because lenders want credit risk models that are stable and accurate over time, Equifax tests and validates the predictiveness of NeuroDecision® Technology compared to models that use an unconstrained, unexplainable neural network algorithm. While the unconstrained model is more accurate initially, it is overfitted to training data so that by the time these models are deployed they are less accurate than the constrained models.

Creditworthiness assessment models that use NeuroDecision® Technology are more predictive than traditional models (such as logistic regression) and provide information, when requested, that is actionable and presented in terms readily understood.

# QUESTIONS

## Definitions

*Question 1 Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?*

We would make a number of observations with respect to the definitions given.

### Artificial intelligence

The given definition: Artificial intelligence (AI) refers to an engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or parameters without explicit programming. AI systems are designed to operate with varying levels of automation.

#### Observations:

- AI is a field of science, and an AI system is an engineered one.
- Rule-based systems fall under the umbrella of AI, which do not necessarily provide predictive outputs and may require explicit programming.

Proposed definition: Artificial Intelligence (AI) is a field of science concerned with building computers and machines that can reason, learn, and act in such a way that would normally require human intelligence or that involves data whose scale exceeds what humans can analyse.

### Machine Learning

The given definition: Machine learning are the patterns derived from training data using machine learning algorithms, which can be applied to new data for prediction or decision-making purposes.

#### Concerns:

- ML is a program or system that learns patterns, and not the patterns itself.

Proposed definition: Machine Learning (ML) is a subfield of AI, a program or system that trains a model and learns patterns from input data to make useful predictions from new or never before seen data. ML gives the computer the ability to learn without explicit programming.

### Large Language Model

The given definition: A large language model (LLM) is a type of generative AI that specialises in the generation of human-like text.

#### Concerns:

- 'human-like text' is inaccurate.

Proposed Definition: Large Language Model (LLM) is a type of generative AI that specialises in the generation of novel combinations of text in the form of natural sounding language.

## Coordinating AI governance

*Question 4 Potential gaps in approaches Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.*

Any proposed AI regulation - voluntary or otherwise - should be cognisant of the multiple existing regimes already in place for data, including the Privacy Principles; Part IIIA of the Privacy Act (credit reporting) and most recently formal Privacy Safeguards introduced as part of the Consumer Data Right. There is a potential for uncertainty or even contradictory provisions arising from an additional set of legislated obligations.

The introduction of any AI governance should avoid the carve outs that characterised the original Privacy Act, which exempted the private sector for the first decade of operation and does not cover state or territory government agencies, nor is mirrored in equivalent state Acts.

We note that the Commonwealth's 2018 AI ethics principles has been followed in 2022 by NSW releasing an AI Assurance Framework applicable to state agencies, using NSW AI Ethics Principles. Most recently a NSW Parliamentary Committee was set up with terms of reference that include consideration of current NSW AI laws and their effectiveness and enforcement.

As a starting point, like the recently announced National Strategy for Identity Resilience, the Commonwealth and States should agree on a single consistent approach to AI regulation, working together on common objectives, standards and practices.

## Transparency

*Question 9 Target areas - Given the importance of transparency across the AI lifecycle, please share your thoughts on:*

- a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?*
- b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.*

In Australia, credit reporting is heavily regulated by a specific section of the Privacy Act, Part IIIA.

Part IIIA captures credit scores as part of "credit reporting derived information"; this in turn can only draw from a restricted list of elements (credit reporting information) set out in the legislation, meaning credit scores cannot consider elements such as a person's income or good history of paying rent on time.

Additionally, the disclosure of credit reporting information, including credit scores, is heavily restricted (e.g. a real estate agent could not request Equifax to disclose to them a rental applicant's credit report or score).

In 2021, legislation was passed to give greater consumer understanding of the algorithms that placed consumers within five score bands (“excellent” “good” “poor” etc). This legislation requires a consumer to be told of the key contributing factors that characterise their score band.

Using patented technology, Equifax pioneered greater transparency, providing consumers with the four contributing factors based on that individual’s actual behaviours. In this way, Equifax ensures transparency and explainability is both understandable and actionable by a consumer.

In this example, the context, use case and stakeholders involved have led to a highly beneficial transparency; the explainability given does not disclose intellectual property, nor undermine the intended purpose of the algorithm or the AI behind it.

## **Increasing trust**

*Question 11 Target areas - What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?*

Building trust is dependent on lifting AI literacy, educating people on AI benefits and risks and putting in place clear processes for disclosures about the AI involved in a transaction.

An important point to make as part of this education is the harm of not using AI systems.

For example, potential bias in credit scores continues to be the subject of ongoing study and testing, but credit risk algorithms are far less likely to produce the discrimination and inaccurate assessment of risk seen in the era of mortgage applicants being subject to in-person interviews.

## **Risk Based Approach**

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

A risk based approach needs to:

- Balance potential harms with the many social and economic benefits promised by AI;
- Target the right use cases, taking into account the likelihood of harm and as well as the severity of harm;
- Have an onus to put in place mechanisms to detect and mitigate unintended consequences and harm;
- Consider the cost of not using AI in terms of forgone benefits.

Risks and impacts also have the potential to change quickly and any governance approach needs to consider how implementation can be done while still allowing for the capacity to change and grow.

While smaller organisations may struggle with compliance, potentially creating barriers to entry, this does not necessarily mean compliance should vary by size of organisation. Much of the standards and rules that surround data and by way of extension, AI, already exist (eg. the Privacy Act) and is therefore unsuitable for the approach used by the Government for financial services Fintechs (an ASIC enforced

enhanced regulatory sandbox).

There are already multiple regimes for data in Australia and a new governance structure brings the risk of overlap, confusion or contradiction. Moreover, because the field is evolving so rapidly, a supervisory or regulatory agency needs to be continuously monitoring and reviewing developments in a way that has not been necessary previously.

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

One of the main limitations is that the harm of not using AI systems is not included. In other words, we recommend a comparative risk-based approach to consider the trade-off between potential impacts and risks of AI versus non-AI systems.

An initial proposition would be that it is preferable to have an imperfect AI system that demonstrates a better performance on crucial life-impacting tasks than a flawed non-AI approach. While more harm is created by not using the AI system, there then needs to be an onus to detect and mitigate unintended AI consequences and harm.

As an example, the factors included in credit decisioning assessments can be complex. An automated system assessing the patterns and biases in the data is capable of a significantly more nuanced outcome than an individual attempting to weigh the same factors.

AI is a space that will continue to change quickly, and the risks and impacts will mirror that rate of change; this is a challenge for any legislative approach, which will not have the ability to keep pace.

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

While Equifax supports a risk based approach, there is ambiguity in how the risk levels are currently defined. Work is needed to elaborate the use cases being captured and further consider how best to define risk levels to target the right use cases and take into account the likelihood of harm and not just the severity of the harm. Another factor to consider is that impact to individuals can vary on spectrum even within a given use case.

A suggestion to complement the elements presented in attachment C could be “AI in the loop assessments” prior to ‘human in the loop’. This can be considered as ‘AI to mitigate AI risks’ or ‘AI to audit AI’ to address ‘scalability’, which is one of the main challenges of ‘human in the loop’ assessments.

For example, one of the risks raised by using ML/AI algorithms is bias or discrimination against certain groups of people that can have life-impacting consequences (e.g., in the assessment of home loan applications). While it would be infeasible for human or manual monitoring to capture bias at scale, AI methodologies like ‘adversarial debiasing’ can potentially mitigate or downscale this issue. The AI outcome could then be passed through the human in the loop assessments for the final review. Some of the patents at Equifax focus on AI to mitigate AI-based risks/issues.

Finally the current definitions for the requirements of explanation for medium and high risk are problematic for both organisation and consumer. Detailed AI outputs create intellectual property risks and may bring a level of detail and technicality that is not actionable or informative to consumers.

