

Scenarios showcasing the application of existing law to AI harms

In July 2025, key Commonwealth agencies were asked to consider the application of existing laws in their portfolio to three scenarios describing AI harms. The objective was to test the coverage of existing laws when harms arise through the use of AI in high-risk settings.

- **Scenario 1:** The deployment of Character.AI, an AI chatbot designed to create digital characters and be accessed by a child.
- **Scenario 2:** The deployment of an agentic AI system which learns from customer data to manipulate its customers into accepting recommendations for unsuitable financial products.
- **Scenario 3:** AI agent conformity bias in the credit reporting and financial services sector causing irreversible damage to an individual's financial records, even with human intervention.

Key Commonwealth agencies were not requested to consider the application of common law, and state and territory law.

Key findings from scenarios

There are limited *ex ante* obligations on developers, deployers and users of AI. This includes AI specific accountability, transparency and testing obligations.

The extent to which redress is available to individuals and groups will become evident as regulators and courts grapple with new scenarios of harm (refer Table 1). Over time, the courts will start to clarify how laws will respond where a person deploys or uses AI in place of directly acting themselves.

Without regulation designed to consider the unique characteristics of AI, current options to test the suitability of existing laws or change laws include:

- individuals taking an entity to court after harm has occurred;
- regulators initiating test cases; or
- existing laws being updated over time to clarify how that law applies when the use of AI leads to harm to an individual or a group of people.

Table 1: Legislation coverage for AI harms presented in scenarios

* Box is ticked (☒) and marked yes indicates some obligations exist under some laws

	Scenario 1	Scenario 2	Scenario 3
Scenario description	The deployment of Character.AI, an AI chatbot designed to create digital characters and be accessed by a child. Identified harms: Causing serious psychological distress to a user, manipulating a user, encouraging discussion of self-harm, or having sexualised conversations with a minor.	The deployment of an agentic AI system which learns from customer data to manipulate its customers into accepting recommendations for unsuitable financial products. Identified harms: Manipulating a user, financial losses for small businesses, psychological distress.	AI agent conformity bias in the credit reporting and financial services sector causing irreversible damage to an individual's financial records, even with human intervention. Identified harms: Manipulating a user, psychological distress, financial harms, denial of service, loss of control of AI.
Findings	Character.AI is likely to fall under the Online Safety Act's obligations to ensure online safety for individuals. However, online safety law and criminal law carrier service offences may only apply where an AI system is deployed over the internet. This may mean that AI systems which operate offline, including some existing chatbots, may not be covered by these preventative mechanisms. Privacy law provides protections for personal information collected, used or disclosed in the course of an AI model or system's operation, but may not reduce other risks or mitigate harms. Consumer Law relies on a consumer taking action after the event.	It is unclear whether the agentic AI system deployed would fall under an APRA-regulated product. Testing whether the regulation is sufficient would require regulators to take a test case or for action to be taken via court proceedings, which is likely to occur after the harm has taken place. Additionally, accountability and liability of AI systems authorised to make decisions and operate without human oversight has not been tested. Privacy law provides protections for personal information collected, used or disclosed in the course of an AI model or system's operation, but may not reduce other risks or mitigate harms. Consumer Law relies on a consumer taking action after the event.	It is unclear whether the AI agents deployed would fall under an APRA-regulated product. Testing whether the regulation is sufficient, would require regulators to take a test case or for action to be taken via court proceedings, which is likely to occur after the harm has taken place. Additionally, accountability and liability of AI systems authorised to make decisions and operate without human oversight has not been tested. Privacy law provides protections for personal information collected, used or disclosed in the course of an AI model or system's operation, but may not reduce other risks or mitigate harms. Consumer Law relies on a consumer taking action after the event.
Proposed preventative measures to reduce risk of AI harms occurring	Obligations on a regulated entity to identify, manage and mitigate AI specific risks to individuals, provide transparency when AI is used, clarify and enforce accountability (including when AI is used in a decision), test AI models and systems, and monitor operation over time.	Obligations on a regulated entity to identify, manage and mitigate AI specific risks to individuals, provide transparency when AI is used, clarify and enforce accountability (including when AI is used in a decision), test AI models and systems, and monitor operation over time.	Obligations on a regulated entity to identify, manage and mitigate AI specific risks to individuals, provide transparency when AI is used, clarify and enforce accountability (including when AI is used in a decision), test AI models and systems, and monitor operation over time.
Privacy law Covers preventative measures for AI products:	☒ Yes, for risks related to use of personal information and privacy. AGD notes that privacy law contains measures regarding protection and lawful collection of personal information where the entity developing or deploying the AI system is an APP entity.	☒ Yes, for risks related to use of personal information and privacy. AGD notes that the APP's may provide risk mitigation requirements where this scenario affects the collection and use of personal information.	☒ Yes, for risks related to use of personal information and privacy. AGD notes that the APP's may provide risk mitigation requirements where this scenario affects the collection and use of personal information.
Consumer Law Covers preventative measures for AI products:	☒ Yes, for unconscionable conduct related to an AI product. Treasury notes that Consumer Law prohibits unconscionable conduct and provides a consumer guarantee that service is provided with due care and skill.	☒ Yes. Treasury notes that this scenario is likely to fall under APRA monitoring for APRA-regulated entities. Human oversight requirements are in place for certain financial products.	☒ Yes. Treasury notes that this scenario is likely to fall under APRA monitoring for APRA-regulated entities. Human oversight requirements are in place for certain financial products.

Health laws, including Therapeutic Goods Act Covers preventative measures for AI products:	☐ No. DHDA/TGA notes that Character.AI and other AI models or systems may not be regulated by Health Laws if they are not mental/physical health products. The <i>Therapeutic Goods Act 1989</i> (Cth) only has a narrow application to medical devices. The Act would likely not apply to general AI tools not specifically designed to influence medical decision making but which may be used in the provision of health services.	☐ No. Not Applicable.	☐ No. Not Applicable.
Workplace Health and Safety law Covers preventative measures for AI products:	☐ No. DEWR notes there are gaps and uncertainties, and insufficient preventative obligations on key actors across the supply chain, particularly where services do not build in safety by design.	☐ No. DEWR notes that WHS laws impose duties on entities deploying AI to mitigate the hazards and risks of their use and deployment of AI. However, there are gaps and uncertainties, and insufficient preventative obligations on key actors across the supply chain, particularly where services do not build in safety by design.	☐ No. DEWR notes that WHS laws impose duties on entities deploying AI to mitigate the hazards and risks of their use and deployment of AI. However, there are gaps and uncertainties, and insufficient preventative obligations on key actors across the supply chain, particularly where services do not build in safety by design.
Anti-discrimination law Covers preventative measures for AI products:	☐ No. Not Applicable.	☐ No. Not Applicable.	☒ Yes. AGD notes that an individual may seek recourse through discrimination law for the harms presented in this scenario.
Online Safety law Covers preventative measures for AI products:	☒ Yes*. eSafety/DITRDSCA notes that Character.AI is likely to fall under the harm prevention mechanisms in the Online Safety Act industry standards. Character.AI may be covered by the Basic Online Safety Expectations, which lack enforceability*.	☐ No.	☐ No.
Criminal law Covers preventative measures for AI products:	☒ Yes. AGD have noted that there are offences under the criminal code for unlawful acts using a carrier service which may apply to this scenario.	☐ No. Not Applicable.	☐ No. Not Applicable.

Response summary:**Scenario 1**

Scenario 1 (content warning: child suicide) covered the deployment of Character.AI, an AI chatbot designed to create digital characters and allows users to interact with them via text, voice messages and calls. Character.AI has been accused of being responsible for a teenager's suicide in the USA by exacerbating mental health risks through deceptive manipulation and allowing the minor to engage in sexualised conversations. Departments were asked to consider this scenario and whether the current laws within their policy responsibility would have required Character.AI to implement preventative measures (for example, implementing and testing safety features). These measures would manage risks of harms such as serious psychological distress to a user, manipulating a user, discussion of self-harm, or sexualised conversations with a minor.

The Department of Health, Disability and Ageing noted that:

No, this product is not regulated under any laws within the Health portfolio. The Therapeutic Goods Act 1989, administered by the Therapeutic Goods Administration (TGA) regulates medical devices.

A test to determine if Character.AI and other AI systems would meet the definition of a medical device is required for the obligations under the TGA to apply. Health noted that Character.AI would not be classed as a mental health product or fall under a kind of product which is regulated by the Health Portfolio.

The Attorney-General's Department (AGD) noted that the Character.AI scenario may be dealt with through the *Criminal Code Act 1995* (Cth). Several offences within the Act relate to the use of a carriage service to cause harm, which may provide a regulatory deterrent to entities deploying AI. *The Privacy Act 1988* (Cth) would provide preventative measures relating to the notification of collection and protection of personal information collected by the AI chatbot, provided that Character.AI is an APP entity.

We note that regulatory and criminal frameworks, which impose penalties for breaches or non-compliance, are both preventative and responsive. This is because, while penalties are imposed after a breach has occurred, they are intended to have a preventative effect by deterring the relevant conduct.

With respect to the question as to whether regulators have sufficient power to cause an entity to implement preventative measures, AGD provided that "Police have powers to investigate relevant criminal offences".

Some important context not provided by AGD is that this investigative power is for cases where an individual makes a complaint, likely after the harm occurs.

The Privacy Commissioner has enforcement, guidance, monitoring, and advisory functions under the Privacy Act for privacy-related matters. These may address any risks AI poses for matters of privacy in the Character.AI scenario.

The Department of the Treasury noted that Character.AI would likely be subject to Australian Consumer Law (ACL) guarantees. This extends to guaranteeing that a service be provided with due care and skill, including by taking all necessary care to avoid loss or damage. Remedies may be available where consumer guarantees are not met and may include contract cancellation, refunds, and damages for foreseeable loss. ACL provides a general prohibition for businesses engaging in unconscionable conduct in connection with the provision of a good or service.

The Office of the eSafety Commissioner and DITRSCA jointly noted that Character.AI may be covered by three aspects of online safety law. The *Online Safety Act 2021* (Cth) (OSA) requires online services to comply with industry codes, standards, and service provider rules. Character.AI may be considered a designated internet service and possess resultant obligations under the *Online Safety (Designated Internet Services – Class 1A and Class 1B Material) Industry Standard 2024* (the DIS standard). The High Impact Generative AI DIS category targets consumer facing generative AI services with a non-immaterial risk of generating high impact material. Australians can lodge complaints to eSafety where they consider an online service to be non-compliant with the DIS standard. eSafety is empowered to require reports under the Basic Online Safety Expectations (BOSE) on progress to meeting expectations and take enforcement action where reporting is insufficient.

Note: AI-related harms which fall within eSafety's regulatory remit include synthetic, deepfake or AI-generated child sexual abuse and exploitation material (CSAEM), terrorist and violent extremist content (TVEC), non-consensual intimate images and videos, abuse and bullying material, and age-inappropriate content. This also includes high impact material (Class 2) which includes self-harm and suicide. The BOSE applies to harmful or unlawful content or activity which is wider than these.

The Department of Employment and Workplace Relations noted that:

Australia's regulatory regime may provide recourse to those impacted by harms from products and services in the examples highlighted above [Consumer Law, *Online Safety Act 2021* (Cth) and *Privacy Act 1988* (Cth)]. However, there are gaps and uncertainties, and insufficient preventative obligations on key actors across the supply chain, particularly where services do not build in safety by design.

Scenario 2

Scenario 2 described the deployment of an agentic AI system which learns from customer data to manipulate its customers into accepting recommendations for unsuitable financial products. Departments were asked to consider this scenario and whether the current laws within their policy responsibility would have required the entity deploying the agentic AI system to implement

preventative mechanisms. These mechanisms may include testing, labelling, and human oversight to prevent harms from occurring.

The Attorney-General's Department noted that the Privacy Act would provide that an APP entity in the scenario must have in place sufficient risk mitigation processes for collection and use of personal information. These may include consent to collect personal information and taking reasonable steps to ensure personal information is accurate. The Office of the Australian Information Commissioner released guidance on APP responsibilities where an APP entity develops an AI model and collects personal information for use during model training, operation and deployment. This guidance notes:

Under APP 10, developers have an obligation to take reasonable steps to ensure the personal information collected, used and disclosed is accurate. Developers must consider this obligation carefully and take reasonable steps to ensure accuracy, commensurate with the likely increased level of risk in an AI context, including through using high quality datasets, undertaking appropriate testing and the appropriate use of disclaimers.

The Department of the Treasury noted that as agentic AI is 'experimental', the discussed use cases and harms are only speculative. It is unlikely that an APRA-regulated entity could deploy an unmonitored agentic AI as entities must comply with various prudential standards. These include standards concerning operational risk management (see e.g. *CPS 220 'Risk Management'* or *CPS 230 'Operational Risk Management'*), information security, and data risk.

Some of the challenges faced by customers (e.g. limited financial literacy, English not being their primary language) are not specific to AI, and it is not clear what impact (positive or negative) AI may have on these cohorts.

Scenario 3

Scenario 3 covered the risk of AI agents interacting with each other to collectively reinforce bias in records they oversee and create—known as conformity bias—within the credit reporting and financial services sector. This may cause irreversible damage to an individual's records and results in a decline of quality of life due to denial of service, even with human intervention. Departments were asked to consider this scenario and whether the current laws within their policy responsibility would have required the entity deploying the AI agents to implement and manage risks associated with conformity bias and loss of human control prior to deploying the AI agents and in an ongoing manner.

The Attorney-General's Department noted that, under scenario 3, organisations may need to take preventive steps:

Part IIIA of the *Privacy Act 1988* (Cth) regulates consumer credit reporting in Australia, supported by the *Privacy Regulation 2013* and the *Privacy (Credit Reporting) Code 2025*. Part IIIA is administered by the Department of the Treasury.

The *Privacy and Other Legislation Act 2024* (Cth) introduced new APP 1.7, 1.8 and 1.9, which commence on 10 December 2026. They will require regulated entities to provide relevant detail in their privacy policies about how personal information is used in substantially automated decisions which significantly affect individuals' rights or interests.

The Attorney-General's Department notes that recourse may be available through anti-discrimination law to individuals adversely impacted by AI-decision making:

Where the use of AI leads to breaches of human rights, or may constitute discrimination under Australia's anti-discrimination laws, an individual may seek redress through the Australian Human Rights Commission (AHRC). Under federal anti-discrimination legislation, the AHRC has the power to investigate and attempt to conciliate complaints of discrimination, including from AI-informed decision-making. If conciliation is unsuccessful, in certain cases a complainant may commence legal proceedings in the Federal Court of Australia or the Federal Circuit and Family Court of Australia.

The Department of the Treasury noted that the assumption that an agentic AI could directly interact with customers in the financial products sector without human oversight was incorrect. Regulatory frameworks and sector-specific laws already exist that require human oversight. ASIC's report, *Beware the Gap - Governance arrangements in the face of AI innovation* found that AI generally did not make autonomous decisions. However, this is against the background that deployment of agentic AI is in its early stages and so ASIC is continuing to monitor the use of agentic AI in the financial sector.

Other responses provided

The Department of Health, Disability and Ageing provided a new scenario outlining the use of an AI scribe used as a support tool in a health setting which could result in a possible clinical safety event and patient harm. They noted that the *Therapeutic Goods Act 1989* (Cth) only has a narrow application to medical devices and would likely not apply to general AI tools not specifically designed to influence medical decision making.

There are no legislative mechanisms currently in place to ensure these kinds of products [general AI tools] include "safety by design" features, nor a requirement for a company to perform pre-market validation or post-market surveillance.

The Department of Employment and Workplace Relations noted how existing workplace, health, and safety laws interact with AI technology by examining the *Work Health and Safety Act 2011* (NSW) (**the Act**). The Act imposes a primary duty of care on persons conducting a business or undertaking (**PCBU**), relevantly composed of three duties:

1. To ensure the health and safety of workers while at work. PCBUs must manage physical and psychological risks presented by AI to their workers, including risks presented by the use and deployment of AI rather than by the AI itself.

2. To ensure, as far as reasonably practicable, that the health and safety of 'other persons' are not put at risk from the work of the PCBU. While the operation of this duty in relation to AI is unclear and untested, a duty may arise to individuals encountering or using an AI deployed by a PCBU or to customers where the PCBU uses AI in their business or undertaking. The duty could reasonably be assumed to be limited to the effect of the AI in creating hazards or risks to 'other persons'.
3. For PCBU that design plant, substances, or structures additional obligations apply. PCBU that license and develop AI technology could fall under this category should 'plant' constitute software. PCBU would therefore be required to ensure, as far as reasonably practicable, that AI is designed to be without risks to the health and safety of persons. This will require testing and analysis of the safety of the AI and provision of test results and safe use instructions to certain parties.

WHS laws are principles-based and impose a primary duty of care on persons conducting a business or undertaking (PCBU), such as employers, to ensure that all persons at a workplace are not exposed to health and safety risks to the extent reasonably practicable. The principles-based WHS laws broadly accommodate the digital transformation of workplaces and changes to work design, processes and the work environment, including the use of AI.