Commissioner Briefing: Character.AI

To	Julie Inman Grant, Commissioner
Cc	s 22 General Manager, Corporate and Strategy s 22, Executive Manager, Strategy, Engagement and Research s 22, Executive Manager Industry Compliance and Enforcement s 22 Executive Manager, Enforcement & Capability Uplift s 22, Manager, Industry Insights and Enablement s 22, Manager, Industry Supervision
From	s 22, Assistant Manager Industry, Insights and Enablement
Cleared by	s 22, Manager, Industry Insights and Enablement
Meeting Details	Tuesday 23rd September 2025, 4:30pm – 5:30pm (Pacific Time) In Person: s 47G, Redwood City, CA 94063, USA
Subject	Meeting with Character.AI
Attachments	NIL

Purpose

- This briefing provides information for your meeting with representatives from Character.AI.

Representatives

	Name and Title
External Attendees	s 47F, Chief Legal Officer & SVP of Global Affairs, Character.AI s 47F, Head of Trust & Safety, Character.AI s 47F, General Counsel, Character.AI
eSafety Attendees	Julie Inman Grant, Commissioner s 47E(c), s 47F, GM Corporate & Strategy s 22, Executive Manager, Industry, Compliance and Enforcement
Note Taker	s 22, Executive Manager, Industry, Compliance and Enforcement

Agenda

1 Introductions

2 eSafety to provide update on key priorities and initiatives relevant to BOSE, Codes and Standards

3 Q/A

Background

- You are scheduled to meet with s 47F (Chief Legal Officer & SVP of Global Affairs), s 47F (Head of Trust & Safety) and s 47F (General Counsel) on 23rd September.

Sensitivities & Risks

- The U.S. Federal Trade Commission (FTC) has launched an inquiry into seven companies, including Character.AI, over the safety and impact of AI chatbots on children and teens.
    - The FTC is examining how these chatbots simulate human-like relationships, what safeguards are in place, and how companies disclose risks to users and parents.
    - Scrutiny of Character.AI has intensified due to cases of teen suicide linked to obsession with a character on Character.AI. Parents of Juliana Peralta have recently filed a wrongful death lawsuit against Character.AI in the U.S. This is the second such lawsuit against Character.AI since the case of Sewell Setzer in 2024.

Talking points and questions

Phase 2 Codes

1. Inform Character.AI that the eSafety Commissioner has registered 9 industry-drafted codes which aim to prevent children accessing or being exposed to age-restricted material, including online pornography, high-impact violence material, self-harm material, and simulated gambling material (the Phase 2 Codes). These Codes will come into effect from 27 December 2025.
    - s 47E(d)

- Request clarification from Character AI on whether group messaging with humans as well as AI is possible on the Character AI service in Australia, noting that this is relevant to the code that applies to them.
- Under the DIS Code, if a high impact generative AI DIS is capable of generating online pornography, high-impact sexually explicit material, self-harm material, high-impact violence material, or violence material, they must perform a risk-assessment.
    i. When the risk assessment is performed, the services with the highest risk of exposing children to this material must implement appropriate age assurance measures to determine if a user can access those services.
    ii. Services with a moderate risk must either implement appropriate age assurance measures or implement robust systems that prevent this material from being generated.
        1. Character.AI has some existing safety tools, including self-declaration-based age assurance, opt-in parental controls (by the child), and modified user experiences for teens (aged 13-18).
- Services with a moderate and high risk of generating age-restricted material must also:
    i. have and enforce clear terms and conditions relating to age-restricted material (Character.AI's community guidelines note that pornographic content, self-harm material and violent material is prohibited.)
    ii. provide tools for users to report, flag or make complaints about age-restricted material
    iii. have sufficient personnel to oversee the safety of the service
    iv. provide easily accessible and clear safety information for end-users
    v. update eSafety about relevant changes to the functionality of the service
    vi. report to eSafety on Code compliance when requested.

BOSE

2. Inform Character.AI that eSafety is empowered through the Act to give providers transparency notices which compel information relating to how providers are implementing the BOSE and publish summaries of this information to publicly hold providers to account for the safety of their services.

3. The BOSE provides specific expectations of providers regarding the use of generative artificial intelligence capabilities. These include:

- Reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of generative artificial intelligence capabilities on the service;

- Reasonable steps to proactively minimise the extent to which generative artificial intelligence capabilities may be used to produce material or facilitate activity that is unlawful or harmful.

4. Highlight that we are focusing on GenAI thematically through 2025-26 for future notices.

5. State that Character.AI received a draft non-periodic notice s 47E(d) The draft notice informs Character.AI that we intend to send it a notice to compel information relating to the safety of the Character.AI service with reference to it being a GenAI chatbot. Character.AI was provided with a list of questions the notice is likely to contain, as well as information regarding how it can make submissions relating to the publication of information it provides.

s 47E(d)

Phase 1 Codes & Standards

6. Inform Character.AI that 8 codes and standards are in effect which require service providers to take steps at a systemic level to reduce the risk that their service will be used to solicit, access, distribute or store 'Class 1A and Class 1B' material, including child sexual exploitation material, pro-terror material and crime and violence material.

   AI chatbot services like Character.AI will be regulated under the Designated Internet Services (DIS) Standard.

7. If the service enables messaging between end-users it is considered a Relevant Electronic Services (RES) under the Online Safety Act and will be regulated by the RES Standard. If a service's purpose is to enable online social interaction between multiple end-users and where it satisfied other elements of the Social Media Service (SMS) definition, the service would be considered a SMS. A service that meets the definition of a RES or SMS cannot be a DIS.

Character.AI should conduct its own assessment of which definition/s it meets under the Online Safety Act, and which codes/standard it is regulated by.]

8. s 47G

9. The DIS Standard contains specific obligations for a 'High impact generative AI DIS'. This category applies where the risk of generating high impact material (X18+ or Restricted Content) is not immaterial. Services in this category must comply with requirements including to:
    - Have and enforce terms of use that prohibit end-users from using the service for illegal and restricted material.
    - Implement systems, processes and technologies that prevent generative AI features from being used to generate outputs that contain child sexual exploitation or pro-terror material.
    - Effectively disrupt and deter the use of the service to create child sexual exploitation material
    - Implement user reporting tools, which must be easily accessible through the service with clear instructions on how to use them, and enable the complainant to specify the harm associated with the material.

Safety by Design

10. Ask about Character.AI's implementation of Safety by Design principles.
    - Character.AI expanded their trust and safety team significantly in 2024, hiring key personnel including a Head of Trust and Safety, a Head of Content Policy and additional safety engineering team members.
    - Character.AI has committed to The Inspired Internet Pledge, created by the Digital Wellness Lab at Boston's Children's Hospital to create a safer healthier internet for young people.
    - Character.AI also works with teen online safety experts at ConnectSafely, who advise on its safety-by-design approach during the development of new features.

11. Ask how Character.AI plans to address young people's safety risks on their platform, especially adverse mental health outcomes.
    - Character.AI currently restricts the platform experience for teens using enhanced safety features, such as:
        - content classifiers designed to reduce exposure to sensitive or suggestive content in the under-18 model.

- monitoring all inputs and blocking content that violates their Terms of Service and community guidelines.
- restricting teen users to a narrow selection of characters. These exclude characters that have been reported by other users.

- Character AI has a Parental Insights feature to help parents and carers monitor teen activity, including:
  - Usage time.
  - Character interactions (top characters that a teen has interacted with and frequency).
  - Engagement patterns (time spent with each character).
- These insights do not include chat content as a privacy protection measure.

12. Ask how Character.AI supports the safety of users in relation to self-harm cases, and if there are foreseeable cases in which safety would be prioritised over privacy, e.g. where it may be reasonably necessary to contact law enforcement.

- Character.AI has implemented pop-up resource triggers when users input phrases related to self-harm or suicide, directing them to the National Suicide Prevention Lifeline.

Affective impact

13. Ask how Character.AI aim to improve, track and assess companion usage to mitigate risks and impacts associated with affective and companionship usage such as influence risk, dependencies, stigmatisation, unhealthy relationships and psychological harm.

- Character.AI offers features like voice calls and messages, which may deepen the sense of realism and emotional connection for users. s 47G
- Character.AI has implemented safety features to mitigate these risks, including:
  - Reminders to users that the AI characters are not real people.
  - Alerts to users after an hour of use.
  - Continuous training of models to comply with policies that prohibit non-consensual sexual content, graphic sexual descriptions, and content promoting self-harm or suicide.

ATTACHMENT A

Biographies

s 47F