

Submission on Deepfake Digital Harm and Exploitation Bill

June 2026

1. The [Classification Office | Te Mana Whakaatu](#) (the Office) welcomes the opportunity to submit on the Deepfake Digital Harm and Exploitation Bill. The Bill would change the definition of “intimate visual recording” in the Crimes Act 1961 and the Harmful Digital Communications Act 2015 to include images that are digitally created or altered to appear to show a person’s likeness.
2. This submission is to support the Social Services and Community Committee’s understanding of how the [Films, Videos, and Publications Classification Act 1993](#) (the Classification Act) applies to digitally created or altered intimate visual recordings (including deepfakes). Our submission outlines relevant parts of the Classification Act, provides examples of content we have classified, and sets out relevant research. We are not making recommendations on the Bill or seeking to appear before the Committee, but the Chief Censor is available to advise the Committee if requested.

About the Classification Office | Te Mana Whakaatu

3. The Office is an independent Crown entity focused on keeping New Zealanders safe from content harm. We are established by the Classification Act to classify publications¹ that may need to be age-restricted² or banned, and to provide information about the classification system and classification decisions. Since 2021, we have supported approved streaming services to rate their content for New Zealand viewers.
4. We prevent exposure to harmful content while upholding the right to freedom of expression and empowering New Zealanders to make informed choices about what they, and their rangatahi and tamariki, watch.

¹ Our functions under the Classification Act limit our classification role to “publications”, which include (among other things) films, games, video on-demand content, books, sound recordings, pictures, newspapers, photographs, prints or writings, electronic or computer files, and copies of images or sounds that have been livestreamed.

² For example, R13, R16, or R18. See our [website](#) for more information about age ratings.

About the Classification Act

5. The Classification Act sets out the legal provisions for classifying publications in New Zealand. We apply these provisions to classify entertainment content (such as films released in cinemas or on DVD) and forensic content submitted by law enforcement agencies and the courts.
6. Our classification function is quasi-judicial. Classifications are legally enforceable, and the determination that a publication is objectionable (illegal) can prove a large part of a criminal charge. We are not responsible for investigations or enforcement of the Classification Act. These are conducted by the Department of Internal Affairs (DIA), the Customs Service, and the Police.
7. In making our decisions, we have regard to the rights affirmed in the New Zealand Bill of Rights Act 1990 (NZBORA). Of particular concern to our decisions is section 14, which affirms “the right to freedom of expression, including the freedom to seek, receive, and impart information and opinions of any kind in any form”³. To give effect to this right, we always apply the least restrictive classification available.
8. All classification decisions are subject to independent review by the Film and Literature Board of Review.
9. The Classification Act can only deal with specific publications and the content of those publications. It does not regulate the websites, wider platform systems, AI tools, algorithms, or applications that enable intimate digital images to be created, distributed and recommended.

Relevant types of content that we have classified

10. The kinds of intimate image or deepfake content that we have classified includes:
 - a. Children’s bodies with adult heads (including known celebrities).
 - b. Adult bodies with children and young person’s heads.
 - c. Iterations on a similar theme (very similar computer-generated images of children, possibly generated from an original image).
 - d. Images taken from social media accounts of young people (NZ victims) and modified to remove clothes (‘nudified’).
 - e. Adults masturbating over non-sexual images of children (‘tribute’ content).

³ <https://www.legislation.govt.nz/act/public/1990/109/en/latest/#DLM225513>

Relevant provisions in the Classification Act

Some intimate digital images can be deemed or determined objectionable

11. Section 3 of the Classification Act sets out the criteria for deeming or determining a publication to be objectionable. We can only classify a publication as objectionable if it deals with matters such as “sex, horror, crime, cruelty, or violence in such a manner that the availability of the publication is likely to be injurious to the public good”⁴. If a publication does not deal with any of these matters, then it cannot be classified as objectionable.
12. Publications will be deemed objectionable under section 3(2) if they promote or support, or tend to promote or support, certain listed activities. The provisions in section 3(2) most relevant to classifying intimate visual recordings are:
 - a. Section 3(2)(a) – promotes or supports the exploitation of children, or young persons, or both, for sexual purposes.
 - b. Section 3(2)(b) – promotes or supports the use of violence or coercion to compel any person to participate in, or submit to, sexual conduct.
13. If a publication is not deemed objectionable under section 3(2), the Office will consider the extent, degree, and manner to which the publication deals with certain harmful content listed in section 3(3). The provision in section 3(3) most relevant to classifying intimate visual recordings is paragraph (b):

‘the extent and degree to which, and the manner in which, the publication exploits the nudity of children, or young persons, or both’.
14. Other section 3 criteria could apply, depending on the subject matter. This includes (but is not limited to) content that promotes or supports sexual conduct with or on the body of a dead person (section 3(2)(c)); content that promotes or supports bestiality (section 3(2)(e)); or content that degrades, dehumanises, or demeans a person (section 3(3)(c)).

The Classification Act applies to digitally created or altered publications in the same way as ‘real’ publications

15. The Classification Act is protective in nature. It requires an assessment of the harm that the availability of a publication poses to the public good. What matters when determining the classification of a publication is therefore the effect of the publication, rather than how it was created. We can’t always tell whether technology was involved in the creation of a publication but, even if we can, this is not relevant to the classification. Neither is the competence with which the technology has been used or the realism of the publication. A badly

⁴ In 2005, this gateway criteria was amended to expand section 3(1) to include nude or partially nude images of children or young persons that are ‘reasonably capable of being regarded as sexual in nature’.

drawn, written, computer-generated, or altered publication can still be classified as objectionable.

Publications do not need to depict ‘real’ people or subjects to be classified as objectionable

16. For the same reason as above, it does not matter whether ‘real’ people or subjects are shown in a publication, or what the ‘real’ age of those shown is. Because the Act is concerned about the effect of a publication, the focus is how the subject is portrayed in the publication and whether that portrayal is harmful to the public good. For example, if the effect of a publication is to promote and support the exploitation of children (regardless of whether real children are shown in the publication), this is considered to be a “corrosive injury”⁵.

The threshold to classify a publication as objectionable is high

17. The threshold for deeming or determining a publication objectionable is very high. This is appropriate given the high penalties for possessing and distributing objectionable publications.
18. To meet the ‘promote and support’ criteria (the ‘tends to’ test), the Court of Appeal has said that, to comply with the approach in *Moonen*⁶, there must be a “real and substantial risk” of the publication promoting or supporting the specified activity.⁷ It is not enough that there be “some possibility” of the publication promoting or supporting the activity.⁸ The Court stated that the best synonym for “tends to” is “be likely”.⁹ So, the test “promotes or supports, or tends to promote or support” can be read as “promotes or supports, or is likely to promote or support”.

The Classification Act criteria does not cover the personal impact of intimate digital images and deepfakes

19. If the subject of a publication is an adult, an intimate visual recording is unlikely to meet the criteria for being objectionable unless there is something else in the publication that meets the Classification Act criteria. For example, unless the sexual violence or coercive aspect of sexual or intimate content is evident within the publication itself, the image would be unlikely to meet the threshold for promoting and supporting ‘the use of violence or coercion to compel any person to participate in, or submit to sexual conduct’ (section 3(2)(b)).

⁵ *Moonen v Film and Literature Board of Review (Moonen II)* [2002] 2 NZLR 754, at para 37.

⁶ *Moonen v Film and Literature Board of Review* [2000] 2 NZLR 9 (CA)

⁷ *Society for the Promotion of Community Standards Inc v Film and Literature Board of Review* [2005] NZCA 176; [2005] 3 NZLR 403 (30 June 2005) at [105].

⁸ At [104].

⁹ At [103].

20. The personal violation that accompanies having one's likeness used in an intimate digital image without one's consent is therefore not covered by the Classification Act criteria.

We must consider relevant evidence when classifying a publication and use our expert judgment

21. While our decisions are based on our expert judgment of what is depicted in a publication, we must take relevant context and evidence into account when making a decision if it is available¹⁰. This can include external research, consultation, or information provided by the submitter or the owner of the publication.
22. In determining whether or not a publication is objectionable under section 3(3), we must consider additional matters set out in section 3(4). These matters include the dominant effect of a publication (section 3(4)(a)), the impact of the medium (section 3(4)(b)), whether a publication has any particular merit, value, or importance (section 3(4)(c)), the intended audience (section 3(4)(d)), or intended purpose (section 3(4)(e)), and any other relevant circumstances (section 3(4)(f)).
23. The Courts have observed that the matters in section 3(3) and 3(4) can be relevant to determining whether a publication 'promotes and supports' the activities listed in section 3(2)¹¹ (even though the Act does not require us to consider the section 3(4) factors if section 3(2) applies). This is because there is an overlap between section 3(2) and 3(3), and the more holistic factors in section 3(4) may be relevant to deemed objectionability under section 3(2)¹². In the context of intimate digital images or deepfakes, the evidence that is likely to be relevant is evidence of the age of the subject.
24. Even if we are provided with evidence that an intimate image or deepfake of an adult has been used in a criminal way, we cannot determine a publication to be objectionable unless the publication itself promotes or encourages those crimes. Explicit sexual content involving adults is usually age-restricted to adults. Making content that appears to be normal pornography objectionable under the Classification Act would not be a reasonable limit on freedom of expression.

¹⁰ *Re Vixen Digital Ltd* [2003] NZAR 418 (HC) at [20] (Durie, J)

¹¹ *Snell v New Zealand Customs Service* [2021] NZHC 1235 at [33] and [35] (Cooke, J)

¹² *Attorney General v Film and Literature of Review v K* [2022] NZHC 319 at [37] (Isac, J)

Our Research

Online Misogyny and Violent Extremism: Understanding the landscape (2024)

25. Following the 2019 March 15 terrorist attack and the Covid pandemic, we began to notice a trend towards an increase in hateful ideologies targeting women in online spaces, including the targeting of high-profile women. This seemed to align with misogynistic themes and male supremacist ideologies we were seeing in documents justifying violent attacks circulated by violent extremists.
26. To understand this trend better, we commissioned Hate and Extremism Insights Aotearoa (HEIA) to conduct a focused literature review on online misogyny, a summary of government and social media responses to this type of content, and a content analysis of New Zealand-based posts on fringe platforms¹³. Additionally, the research team at the Classification Office reviewed research reports and sought insights from different government departments, Crown entities, and NGOs.
27. We produced both an [online resource](#)¹⁴ and a [summary report](#)¹⁵ based on the literature review. The key findings were that:
 - a) Women and girls are more likely to encounter gender-based and severe forms of abuse online. This adversely affects their mental health, and these harms are amplified for women and girls with intersecting identities.
 - b) Algorithms amplify misogynistic and extremist content and create pathways for vulnerable individuals to be exposed to more extreme ideologies. There is evidence¹⁶ to suggest that algorithmic bias may amplify misogynistic or abusive content on some online platforms.
 - c) Emerging categories of content and technology, such as AI-generated deepfakes, pose challenges for online safety¹⁷. Deepfakes can re-victimise women and girls by circulating manipulated, non-consensual content, leading to public harassment and significant emotional harm.

¹³ [Case study: content analysis of fringe platforms | Classification Office](#)

¹⁴ [Online Misogyny and Violent Extremism – Online Resource | Classification Office](#)

¹⁵ Te Mana Whakaatu Classification Office. (2024). [Online misogyny and violent extremism: Understanding the landscape \(Summary Report\)](#). Wellington, NZ: Te Mana Whakaatu Classification Office

¹⁶ Rathee, S., Banker, S., Mishra, A. & Mishra, H. (2023). [Algorithms propagate gender bias in the marketplace—with consumers’ cooperation](#). Journal of Consumer Psychology and Binns, R., Veale, M., Van Kleek, M. & Shadbolt, N. (2017). [Like trainer, like bot?](#) Inheritance of bias in algorithmic content moderation. Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9 (pp. 405–415). Springer International Publishing.

¹⁷ [AI-powered tools and the creation of deepfakes | Classification Office](#), <https://www.classificationoffice.govt.nz/resources/research/online-misogyny-and-violent-extremism-index/technology-and-online-platforms/ai-powered-tools-and-the-creation-of-deepfakes/>

28. We have provided more detail on the specific findings from our research on the issue of deepfakes and technology-assisted gender-based violence in [Appendix A](#).

Resources

29. Findings from our consultations with young people¹⁸ and our survey about New Zealanders' experiences and exposure to extreme or illegal content¹⁹ revealed a limited understanding among the NZ public about the line for illegal content. In response, we have created information pages in plain English on our website to explain different categories of illegal content and new technology. These include:

- a. [Banned or illegal content](#)²⁰
- b. [Child Sexual Exploitation Material](#)²¹
- c. [Sexual violence content](#)²²
- d. [What is Grok?](#)²³

30. We created a specific resource called [AI-generated content and when is it illegal in New Zealand?](#)²⁴ to ensure that people are aware that creating objectionable content, including through AI tools, is still a serious offence in New Zealand law, and carries significant penalties including periods up to 14 years' imprisonment –see [Appendix B](#) for the full text of this resource.

Final comments

31. We appreciate the opportunity to submit on this Bill. We are happy to provide further information on our research, this submission or any other questions you may have.

¹⁸ [Content that Crosses the Line: Conversations with young people about extremely harmful content online | Classification Office](#) <https://www.classificationoffice.govt.nz/resources/research/content-that-crosses-the-line/>

¹⁹ [Online Exposure: Experiences of Extreme or Illegal Content in Aotearoa | Classification Office](#), <https://www.classificationoffice.govt.nz/resources/research/online-exposure-experiences-of-extreme-or-illegal-content-in-aotearoa/>

²⁰ [What is Objectionable Content? | Classification Office](#) <https://www.classificationoffice.govt.nz/resources/enforcement-officials/banned-or-illegal-content/>

²¹ [Child Sexual Exploitation Material \(CSEM\) content | Te Mana Whakaatu Classification Office | Classification Office](#) <https://www.classificationoffice.govt.nz/classification-info/what-is-harmful-and-illegal-content/child-sexual-exploitation-material-csem/>

²² [Sexual violence content | Te Mana Whakaatu Classification Office | Classification Office](#) <https://www.classificationoffice.govt.nz/classification-info/what-is-harmful-and-illegal-content/sexual-violence-content/>

²³ [What is Grok? Your questions answered | Classification Office | Classification Office](#) <https://www.classificationoffice.govt.nz/resources/items/what-is-grok/>

²⁴ [AI-generated content and when is it illegal in New Zealand? | Classification Office](#) <https://www.classificationoffice.govt.nz/classification-info/what-we-classify/ai-generated-content/>

Appendix A:

Research on deepfakes and technology-associated gender-based violence

Deepfake technology is used for technology-facilitated gender-based violence

32. Deepfake technology is increasingly being used as a form of technology-facilitated gender-based violence. This includes the creation and circulation of non-consensual synthetic intimate imagery, sexualised deepfakes, and manipulated images, which are used to humiliate, harass, threaten, or coerce victims.
33. In her [2023 paper](#)²⁵ on technology-facilitated gender-based violence, hate speech, and terrorism, Esli Chan notes that internet networks can diffuse user responsibility and allow groups to launch co-ordinated online attacks against individuals and groups, including through hateful messages, images, and posts. These forms of attack can include the creation and dissemination of deepfakes.
34. Recent international research also shows that non-consensual synthetic intimate imagery is not a niche or hypothetical harm. [A 2024 study](#)²⁶ of more than 16,000 respondents across 10 countries found that deepfake pornography behaviours were considered harmful by respondents, despite relatively low public awareness. The study identified experiences of both victimisation and perpetration. The authors concluded that legislation alone is unlikely to be sufficient to prevent harm, and that legal responses should be supported by prevention, education, effective platform policies, detection and blocking tools, and practical removal and redress pathways.

Accessibility to deepfake tools increases the potential for gender-based violence

35. Tools to make deepfakes are easily accessible. [2025 empirical research](#)²⁷ on downloadable deepfake model variants found almost 35,000 publicly downloadable deepfake model variants across two repositories, with nearly 15 million downloads since November 2022. The models targeted a range of

²⁵ Chan, E. (2023). [Technology-Facilitated Gender-Based Violence, Hate Speech, and Terrorism: A Risk Assessment on the Rise of the Incel Rebellion in Canada](#). *Violence Against Women*, 29(9), 1687-1718.

²⁶ 16 Umbach, R., Henry, N., Beard, G. F., & Berryessa, C. (2024). Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3613904.364238>
[Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries | Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems](#)

²⁷ Hawkins, W., Russell, C., & Mittelstadt, B. (2025). Deepfakes on Demand: The rise of accessible non-consensual deepfake image generators. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*.
[\[2505.03859\] Deepfakes on Demand: the rise of accessible non-consensual deepfake image generators](#)

individuals, from global celebrities to lower-profile Instagram users, with 96% of the identified models targeting women.

36. Social media enables dissemination of deepfake content of high-profile women, indicating that deepfake abuse is not just about individual misuse but platform design. In 2024, fake sexually explicit images of Taylor Swift proliferated on X, leading the platform to temporarily block searches for her name²⁸. In a more recent example, British MP Jess Asato has filed proceedings against xAI, alleging that Grok was used to create fake sexualised images of her²⁹.
37. The harm from deepfake technology does not just effect high-profile women overseas. The Ministry for Women has highlighted online harm as a barrier to women's public participation and leadership in Aotearoa New Zealand. [Their 2026 Free to Lead Long-Term insights Briefing](#) notes that New Zealand women in public and leadership roles, including MPs, experience personal, threatening, sexualised, and persistent abuse online³⁰.

Rising use of deepfake technology to sexually exploit children and young people

38. Deepfake and AI-generated sexual content also raises specific concerns for children and young people. UNICEF has warned, in [their 2026 issue brief](#)³¹, that AI-generated CSAM includes peer-to-peer "deepfake" nudes that often disproportionately target girls. [The Internet Watch Foundation](#)'s³² most recent report has also highlighted a significant increase in AI-generated child sexual abuse material. Girls were overwhelmingly represented in illegal AI-generated images assessed by the organisation.

²⁸ Reuters. (2024). Taylor Swift searches blocked on X after fake explicit images spread. [Taylor Swift searches blocked on X after fake explicit images spread | Reuters](#)

²⁹ Reuters. (2026). British lawmaker sues Musk's xAI over sexualised Grok images.

²⁰ Ministry for Women. *Free to Lead: Long-term Insights Briefing*.

[British lawmaker sues Musk's xAI over sexualised Grok images | Reuters](#)

³⁰ Ministry for Women. *Free to Lead: Long-term Insights Briefing*. [Free to Lead Long-term Insights Briefing \(LTIB\) | Ministry for Women](#)

³¹ UNICEF. (2026). *Artificial Intelligence and Child Sexual Abuse and Exploitation*. [Artificial Intelligence and Child Sexual Abuse and Exploitation | UNICEF](#)

³² Internet Watch Foundation. (2026). *Harm without limits: AI child sexual abuse material through generative AI*. [iwf.org.uk/media/hl1nvditi/iwf-ai-csam-report-2026.pdf](#)

Appendix B:

AI-Generated Content: When Is It Illegal in New Zealand?

Artificial intelligence (AI) tools can now create images, videos, audio, and written material that may look highly realistic or completely synthetic. It appears across social media, memes, advertising, gaming, entertainment, and everyday online interactions. Some AI might look like it involves real people – even if it doesn't. Apps, including deepfake or so-called “nudify” tools, make it easy for anyone to generate content that can be harmful, abusive, or illegal.

A common misunderstanding is that AI-generated or “fake” content cannot be illegal because it isn't real. In New Zealand, it can.

Under the [Films, Videos, and Publications Classification Act 1993](#), AI-generated content is treated the same as any other content. What matters is what the content shows, not how it was made. Even if something is fictional, computer-generated, or created as a joke, it can still be illegal.

This page explains how the classification law applies to AI-generated content, what types of AI content can be illegal, why this matters, and what to do if you come across it.

Important: This information is provided to help you understand how the law applies to AI-generated content. It is not legal advice.

What is AI-generated content?

AI-generated content is material created using artificial intelligence tools. This can include:

- Images or videos created from text prompts
- Deepfake videos or audio
- Edited or altered photos
- Synthetic or fictional characters
- AI-generated voices or music
- Written content produced by AI tools

This content may look real even though it has been created by a person using a computer program or app.

Can AI-generated content be illegal in New Zealand?

Yes. AI-generated content can be illegal under New Zealand law.

New Zealand has a high threshold for what is considered illegal content. The Classification Act uses the term “objectionable” to describe content that is so harmful it is illegal to make, possess, or share.

Examples of objectionable content include:

- Child sexual exploitation material (CSEM)
- Sexual violence involving children or young people
- Bestiality
- Extreme violence
- Terrorist or violent extremist content

The law does not require the people or animals depicted to be real. If AI-generated or fictional material meets the legal tests for being objectionable, it is treated the same as real-world depictions.

What does the law say about AI-generated content?

The Films, Videos, and Publications Classification Act 1993 applies to all publications, regardless of how they are made or distributed.

This means:

- It does not matter if the content is AI-generated, animated, drawn, or computer-generated
- It does not matter if it was made using an app, AI model, filter, or editing tool
- It does not matter if it was intended as a joke, meme, experiment, or test
- If the content itself meets the legal criteria for being objectionable, it is illegal to make, possess, or share it

What types of AI content does the law apply to?

The law applies equally to all of the following:

- Real footage that has been manipulated
- Deepfakes
- AI-generated images
- CGI or animated content
- Fictional or synthetic characters

With the growing availability of AI tools, people are increasingly experimenting with creating content. It’s important to understand that intent does not make illegal content

legal. Even if content is created “just to see what the app can do”, it may still break the law.

Why is some AI content harmful?

Illegal content causes real harm, even when it is AI-generated or fictional.

For example:

- AI-generated sexual content involving children normalises abuse and exploitation
- Deepfake sexual images can humiliate, harass, or blackmail people
- Violent extremist content can promote or support real-world violence

Our classification system exists to protect people – especially children and young people – while still respecting freedom of expression.

Where might people come across AI-generated illegal content?

People can encounter AI-generated illegal content in many ways, including:

- Social media feeds or “For You” pages
- Group chats or private messages
- Apps that claim to “undress”, alter, or sexualise images
- Search engine results or websites
- Online forums or gaming platforms
- Someone showing them content offline
- Deliberate searching or accidental exposure

Some people create or share this material believing it is harmless because “it’s fake”. This belief is incorrect under New Zealand law.

Is it illegal to possess or share AI-generated illegal content?

Yes. If AI-generated content is classified as objectionable:

- Making it is illegal
- Possessing it is illegal
- Sharing, sending, or uploading it is illegal

Serious penalties can apply for offences under the Classification Act.

What should I do if I come across AI-generated illegal content?

If you come across content you think may be illegal:

- Do not share it with others
- Do not save or forward it
- Report it as soon as possible

Where to report AI-generated illegal content

You can report illegal or harmful content through the appropriate channels, depending on where you found it:

- Online platforms: Use in-platform reporting tools on social media sites, messaging apps, gaming platforms, or websites
- Netsafe: For help with harmful online content, scams, or abuse
- New Zealand Police: For content involving serious criminal harm

[How to Report Harmful or Illegal Online Content in New Zealand](#)

What if I accidentally made AI-generated illegal content?

If you believe you may have accidentally created illegal AI-generated content:

- Stop using or sharing the content immediately
- Delete it from your devices and accounts where possible
- Do not upload or distribute it
- Seek legal advice if you are unsure about what action to take

About our role and how the law works

The [Films, Videos, and Publications Classification Act 1993](#) balances freedom of expression with protecting people from harm – especially children and young people.

Every piece of content we see is different. To decide whether something is illegal, we must consider each example on its own merits and carefully apply the Act and the principles of fairness and natural justice. There are also serious penalties for offences under the Act. This is why decisions about whether content is illegal are made by trained experts who follow a very transparent and fair process, guided by the Act, before making a decision.

You can read more about our [classification process here](#).

If you are unsure about AI-generated content, it is safer to not create, keep, or share it. Seek guidance if you are unsure what to do.

Further reading

[What we classify](#)

[What is Objectionable Content](#)

[How to Report Harmful or Illegal Online Content in New Zealand](#)

[Helplines and Support Services in New Zealand](#)