

Submission on Deepfake Digital Harm and Exploitation Bill

June 2026

Introduction

The New Zealand Psychological Society Rōpū Mātai Hinengaro Aotearoa (NZPsS, “the Society”) is the premier professional and scientific association for psychologists in Aotearoa/ New Zealand and is committed to supporting quality practice, education and research in psychology. NZPsS represents over 2300 members and students and encompasses a broad range of practice in psychology in Aotearoa/New Zealand, including clinical, counselling, health, forensic, community, organisational and educational. As an organisation we have a strong commitment to Te Tiriti o Waitangi.

NZPsS supports this Bill and the provisions in it as we recognise the significant psychological harm that can arise from the production and distribution of deepfakes.

By explicitly expanding the definition of ‘intimate visual recordings’ to include created or synthesised deepfakes, it appropriately brings reputationally harmful content (e.g., fake sexual imagery) under existing criminal and digital harm frameworks. This is important because reputational damage and associated distress are well-established consequences of image-based and online abuse (Umar, 2026).

We also support this Bill for the four reasons outlined below.

Deepfake digital harm affects the most vulnerable members of society

The use of deepfake tools, especially ‘nudify’ apps, is a form of sexual abuse that is often used against the most vulnerable members of our society. They can be exploited to generate non-consensual synthetic explicit images, including of children and young people, often with devastating impact (Williams, 2025; eSafety Commissioner, 2025).

In a UNICEF, ECPAT and INTERPOL study across 11 countries, at least 1.2 million children disclosed having had their images manipulated into sexually explicit deepfakes in the past

year. In some countries, this represents one in 25 children – the equivalent of one child in a typical classroom (UNICEF, 2026).

It causes significant psychological harm to victims and others

While evidence is limited, there are some studies that provide empirical evidence on the harms of deepfake images.

Unwanted sharing of intimate images, whether they are created by AI or not, causes psychological harm to the victim, and often those around them. The adverse psychological effects of AI-generated intimate images can be similar to other well-documented online harm, such as rumours, the posting of hurtful comments, and the sharing of embarrassing and private pictures or videos (Flynn, Powell, Eaton, & Scott, 2025).

Deepfake nudes violate an individual's right to bodily autonomy – the control over one's own body without interference. Victims experience a severe invasion of privacy and may feel a loss of control over their bodies, as their likeness is manipulated without consent. This can lead to shame, anxiety, and a decreased sense of self-worth. Fear of social ostracism can also contribute to anxiety, depression, and, in extreme cases, suicidal thoughts. The effects of this harassment can be long-lasting, impacting trust, social relationships, and self-esteem well into adulthood. Coming forward risks further exposing victims to shame and victim-blaming, while the perpetrators remain invisible (Patrone & Viola, 2026; Rutledge, 2024).

Estimates suggest that 96% of existing online deepfakes are pornographic in nature [19], it is therefore of note that there were no empirical studies quantifying the effects of deepfake porn exposure in the scoping review (though our search terms did not seek these studies out). Deepfake pornography can potentially pose a threat to the depicted person's reputation and mental health through its dissemination and subsequent distress, similar to how non-consensual spread of intimate images online can cause emotional distress (Ching et al, 2026).

Negative impacts of consuming AI-generated sexual content or synthetic pornography in one descriptive study were described as:

- addiction and dependency risks, lack of control over viewing
- lowered interest in real sexual interactions due to the combination of customisation and instant gratification
- distorted expectations of real sexual interactions and romantic and/or sexual relationships
- harm to body image of viewers

- exploitation of women, people of color, and children in this content. These identities and ages are disproportionately featured in synthetic pornography (Kim & Banks, 2024).

Research has also shown that sharing intimate images without consent causes psychological harm comparable to direct sexual abuse (Naezer & van Oosterhout, 2021).

A review presenting a general overview on the research on harms of deepfakes to human mind and behavior, showing that deepfakes are associated with widespread concerns, negative consequences of successful deception, and distress caused by targeted victimisation (Diel et al, 2025; American Academy of Paediatrics, *n.d.*).

There is an increase in prevalence of deepfakes

Netsafe has reported that use of deepfakes in Aotearoa New Zealand is increasing rapidly, so this problem will only get worse. It has reported that the number of cases of deepfakes used for image-based sexual abuse reported to Netsafe doubled in one year and is set to do so again this year. Reports of image-based sexual abuse have more than tripled from 1000 to 3500 between 2020 and 2025. Most age brackets saw an increase – particularly between 22-40-year-olds, but also among 13–17-year-olds and those over 64 (McCallum, 2026).

Regulation is a necessary part of the solution

It is a social injustice to leave those most at risk of being harmed by deepfakes to have to protect themselves from them.

UNICEF has called on all governments expand definitions of child sexual abuse material to include AI-generated content, and criminalise its creation, procurement, possession, and distribution (UNICEF, 2026).

Privacy regulators in more than 60 countries have called for legislation against deepfakes (Jahangir, 2026) and many, including the European Union, US, UK, Australia, Japan, and China, have already done so.

Conclusion

We support this bill in it's purpose to expand the definition of 'intimate visual recordings' to include created or synthesised deepfakes. Bringing these types of images under existing criminal and digital harm frameworks recognises the significant harm, including psychological impacts that deepfakes can cause.

References

American Academy of Pediatrics (*n.d.*). The impact of deepfakes, synthetic pornography, & virtual child sexual abuse material. *Center of Excellence on Social Media and Youth Mental Health*, US.

Ching D, Twomey J, Aylett MP, Quayle M, Linehan C, Murphy G. 2025. Can deepfakes manipulate us? Assessing the evidence via a critical scoping review. *PLoS One*. <https://doi.org/10.1371/journal.pone.0320124>

Diel, A., Lalgı, T., Mellis, F. et al. The harm of deepfakes: a scoping review of deepfakes' negative effects on human mind and behavior. *AI & Soc.* <https://doi.org/10.1007/s00146-025-02774-0>

eSafety Commissioner. (2025). Deepfake damage in schools: How AI-generated abuse is disrupting students, families and school communities. *eSafety Commissioner*, Australia.

Flynn, A., Powell, A., Eaton, A., & Scott, A. J. (2025). Sexualized deepfake abuse: Perpetrator and victim perspectives on the motivations and forms of non-consensually created and shared sexualized deepfake imagery. *Journal of Interpersonal Violence*. Advance online publication. <https://doi.org/10.1177/08862605251368834>

Jahangir, R. (2026). Privacy regulators in 61 countries back enforcement against AI deepfakes. *Tech Policy Press*, US.

Kim, S., & Banks, J. (2024). Expanding experiences and anxieties: gender-identity and sexual-orientation differences in attitudes toward synthetic pornography. *Porn Studies*, 1–19. <https://doi.org/10.1080/23268743.2024.2400945>

McCallum H. (2026). Netsafe sees surge in reports of sexual deepfake abuse. *Newsroom*, New Zealand.

Naezer, M., & van Oosterhout, L. (2021). Only sluts love sexting: youth, sexual norms and non-consensual sharing of digital sexual images. *Journal of Gender Studies*, 30(1), 79–90. <https://doi.org/10.1080/09589236.2020.1799767>

Patrone, F., Viola, M. Non-consensual sexual deepfakes as direct personal harm. *Philos. Technol.* 39, 94 (2026). <https://doi.org/10.1007/s13347-026-01107-3>

Rutledge, P. (2024). Teens and deepfake nudes: Digital bullying's new frontier. *Psychology Today*, US.

UNICEF. (2026). 'Deepfake abuse is abuse': Statement by UNICEF on AI-generated sexualised images of children. *UNICEF*.

Umar, M. (2026). Digital trauma: Deepfake victimisation and AI-generated violence. *The Lancet Psychiatry*, 13(2), 89–90. [https://doi.org/10.1016/S2215-0366\(25\)00275-5](https://doi.org/10.1016/S2215-0366(25)00275-5)

Williams, K. (2025). “There are no limits!”: AI undressing apps and the normalization of nonconsensual intimate deepfakes. *Violence Against Women*. Advance online publication. <https://doi.org/10.1177/10778012251397966>