AI Tech Abuse: Harms, Limits & Possibilities

PROFESSOR NICOLA HENRY

SOCIAL EQUITY RESEARCH CENTRE

RMIT UNIVERSITY

4 February 2026

Acknowledgement of Country

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

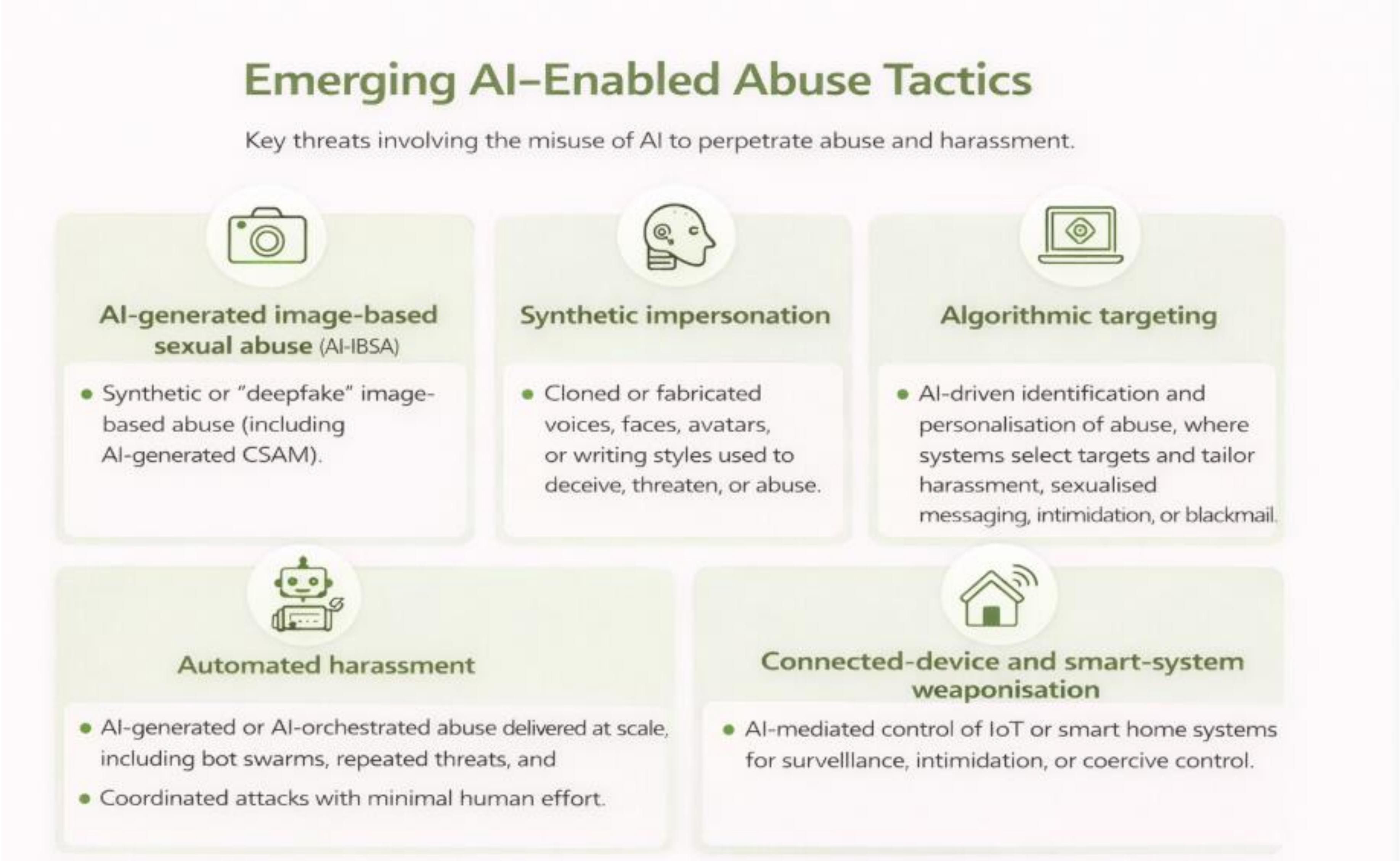
RMIT University respectfully acknowledges their Ancestors and Elders, past and present.

RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.

Artwork ‘Luwaytini’ by Mark Cleaver, Palawa

This presentation will cover topics on intimate images and online abuse.

- Some material may be confronting or distressing.
- Feel free to take a break if you need.
- Visit Tech Safety Canada for more information: https://techsafety.ca/resources/toolkits/image-based-abuse-and-the-non-consensual-distribution-of-intimate-images



Emerging AI–Enabled Abuse Tactics

Key threats involving the misuse of AI to perpetrate abuse and harassment.

AI-generated image-based sexual abuse (AI-IBSA)

- Synthetic or "deepfake" image-based abuse (including AI-generated CSAM).

Synthetic impersonation

- Cloned or fabricated voices, faces, avatars, or writing styles used to deceive, threaten, or abuse.

Algorithmic targeting

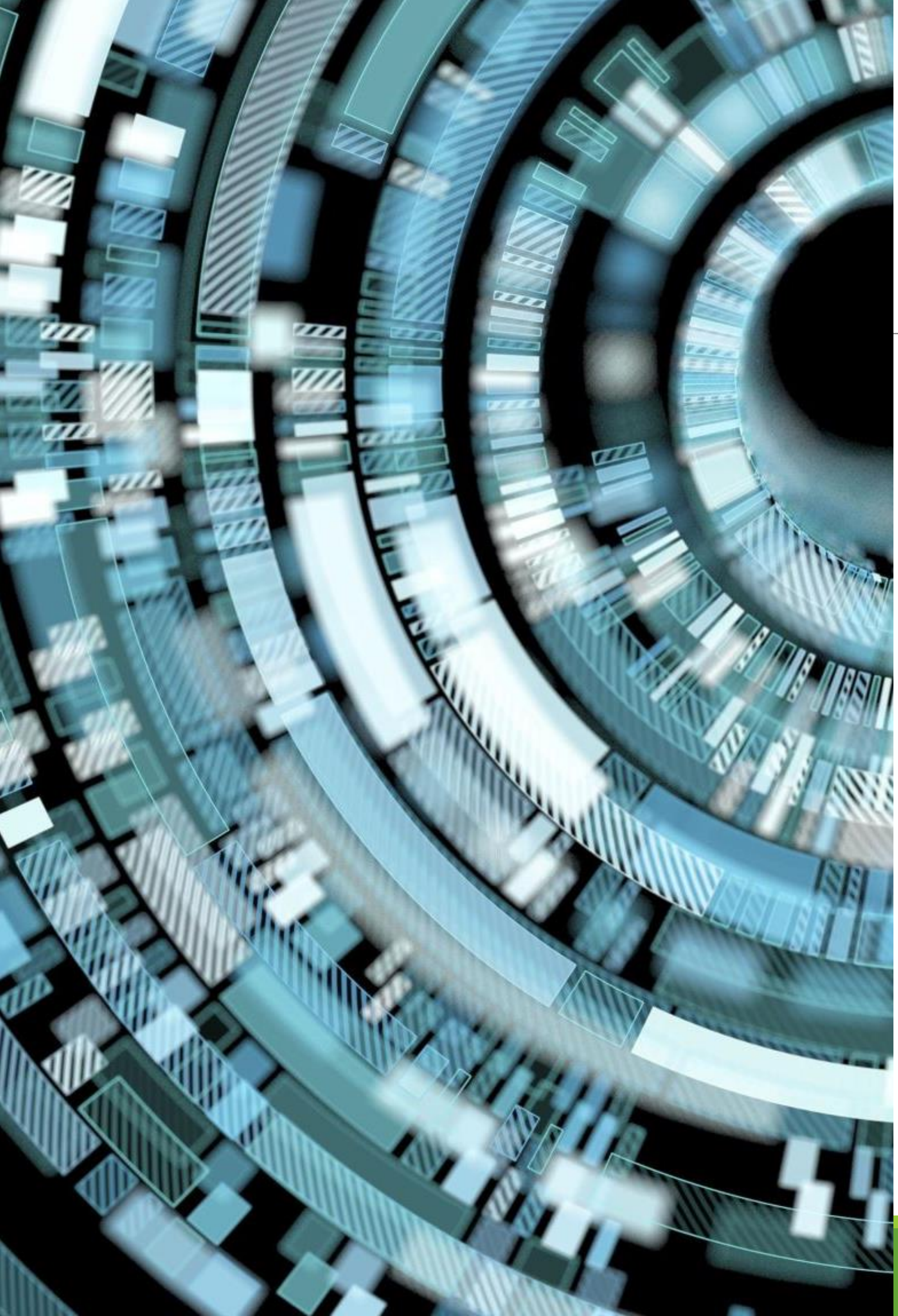
- AI-driven identification and personalisation of abuse, where systems select targets and tailor harassment, sexualised messaging, intimidation, or blackmail.

Automated harassment

- AI-generated or AI-orchestrated abuse delivered at scale, including bot swarms, repeated threats, and
- Coordinated attacks with minimal human effort.

Connected-device and smart-system weaponisation

- AI-mediated control of IoT or smart home systems for survelllance, intimidation, or coercive control.



AI-generated intimate image abuse ("deepfake pornography") (AI-IBSA)

"Deepfakes" refers to the creation of fake but realistic-looking photos or videos using artificial intelligence.

AI-IBSA: nonconsensual creation or sharing of synthetic nude or sexual images, including threats to share images.

Nudify tools and chatbots

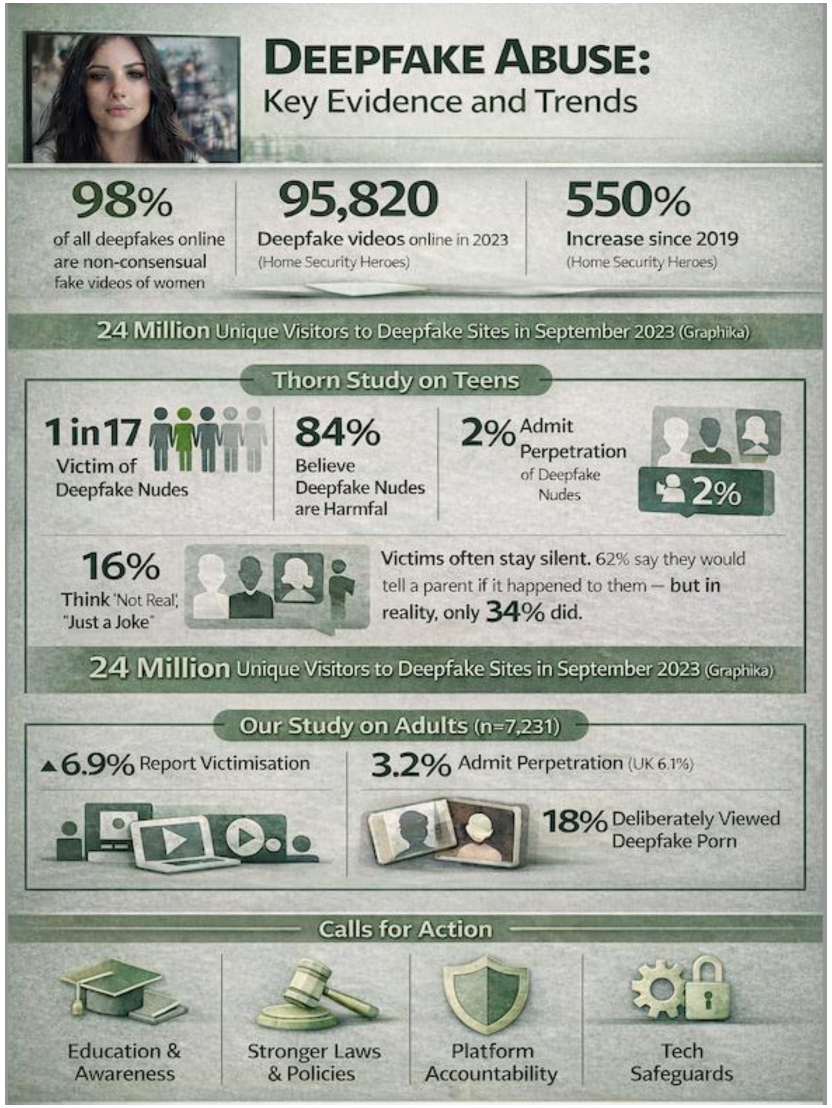
AI-powered apps and websites (including on the Dark Web) that digitally remove clothing from existing images, generating nude or explicit synthetic imagery at a user's request.

Deepfake generators

Deep learning systems that manipulate existing photos, videos, or audio to depict a person doing or saying things they never did — including face swapping, voice cloning, and altering body features or movements.

Generative AI (GenAI)

Tools that create entirely new synthetic sexual imagery from scratch using text prompts or reference images, mimicking the likeness of real people — even when no original photo or video exists in that context.



DEEPFAKE ABUSE:

Key Evidence and Trends

98% of all deepfakes online are non-consensual fake videos of women

95,820 Deepfake videos online in 2023 (Home Security Heroes)

550% Increase since 2019 (Home Security Heroes)

24 Million Unique Visitors to Deepfake Sites in September 2023 (Graphika)

Thorn Study on Teens

1 in 17 Victim of Deepfake Nudes

84% Believe Deepfake Nudes are Harmfal

2% Admit Perpetration of Deepfake Nudes

2%

16% Think 'Not Real', "Just a Joke"

Victims often stay silent. 62% say they would tell a parent if it happened to them – **but in reality, only 34% did.**

24 Million Unique Visitors to Deepfake Sites in September 2023 (Graphika)

Our Study on Adults (n=7,231)

▲**6.9%** Report Victimisation

3.2% Admit Perpetration (UK 6.1%)

18% Deliberately Viewed Deepfake Porn

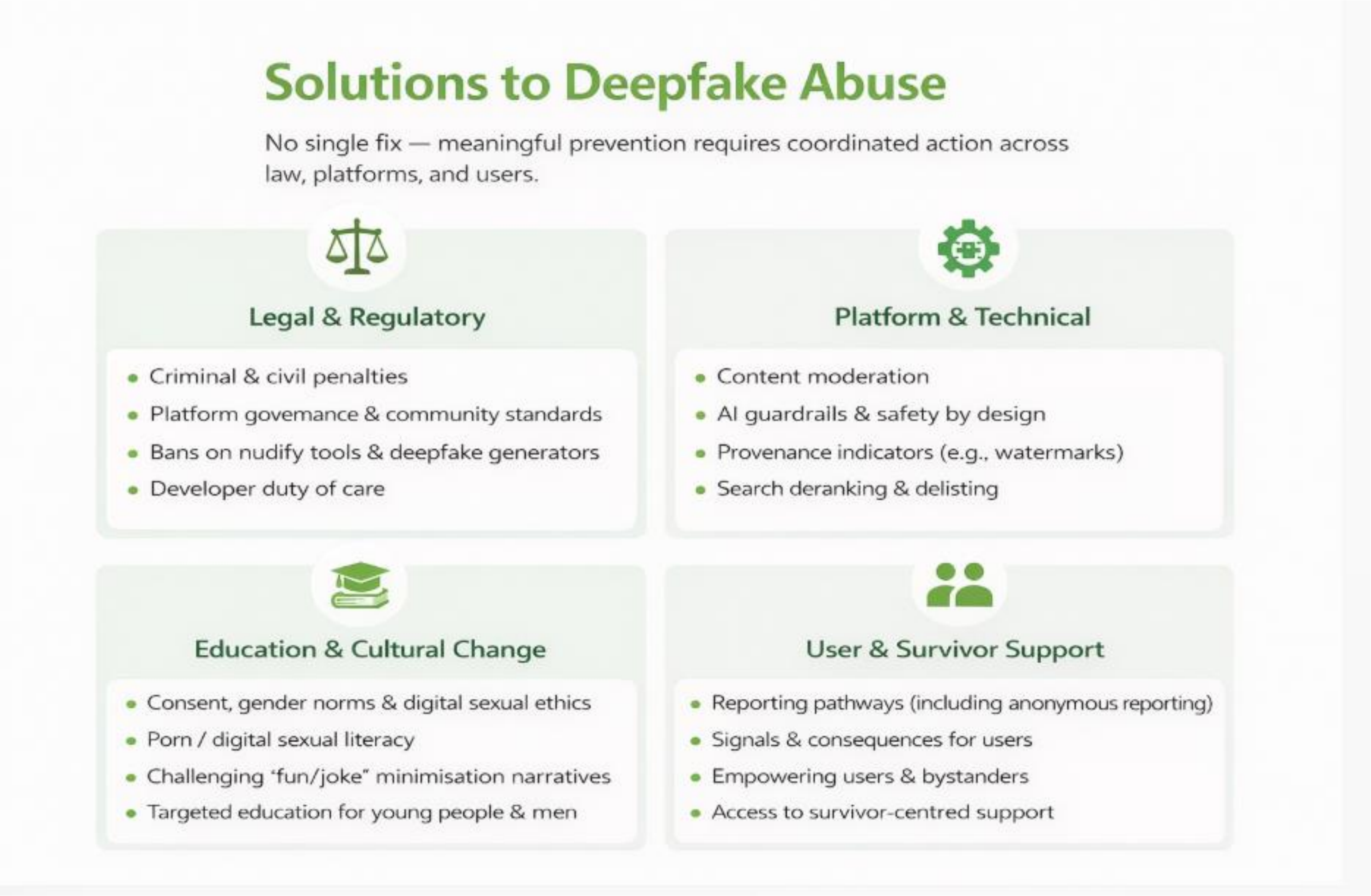
Calls for Action

Education & Awareness

Stronger Laws & Policies

Platform Accountability

Tech Safeguards



Solutions to Deepfake Abuse

No single fix — meaningful prevention requires coordinated action across law, platforms, and users.

Legal & Regulatory

- Criminal & civil penalties
- Platform govemance & community standards
- Bans on nudify tools & deepfake generators
- Developer duty of care

Platform & Technical

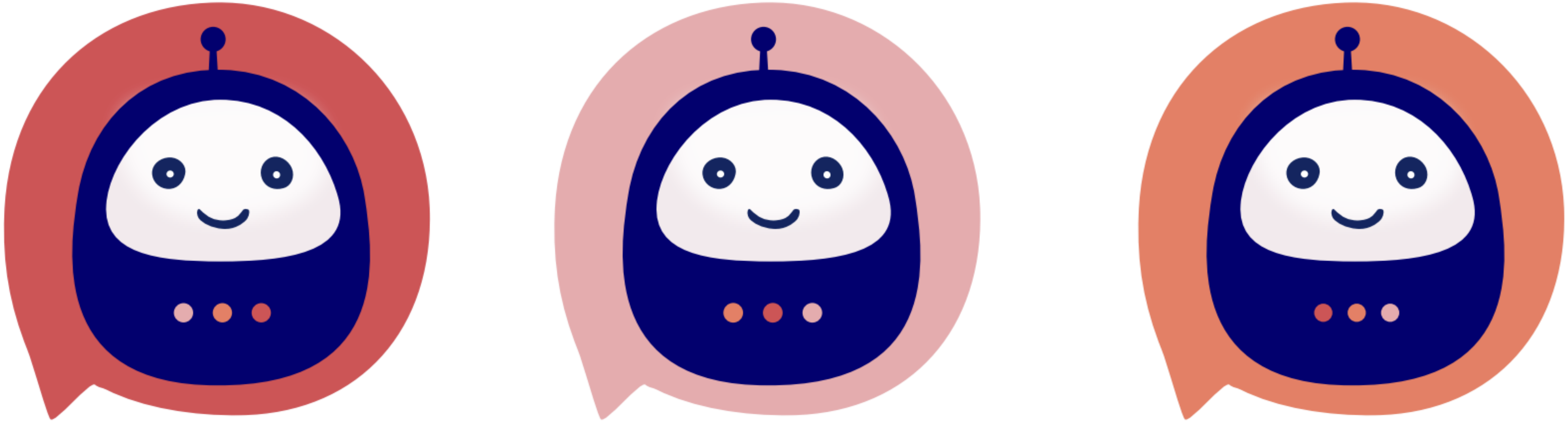
- Content moderation
- AI guardrails & safety by design
- Provenance indicators (e.g., watermarks)
- Search deranking & delisting

Education & Cultural Change

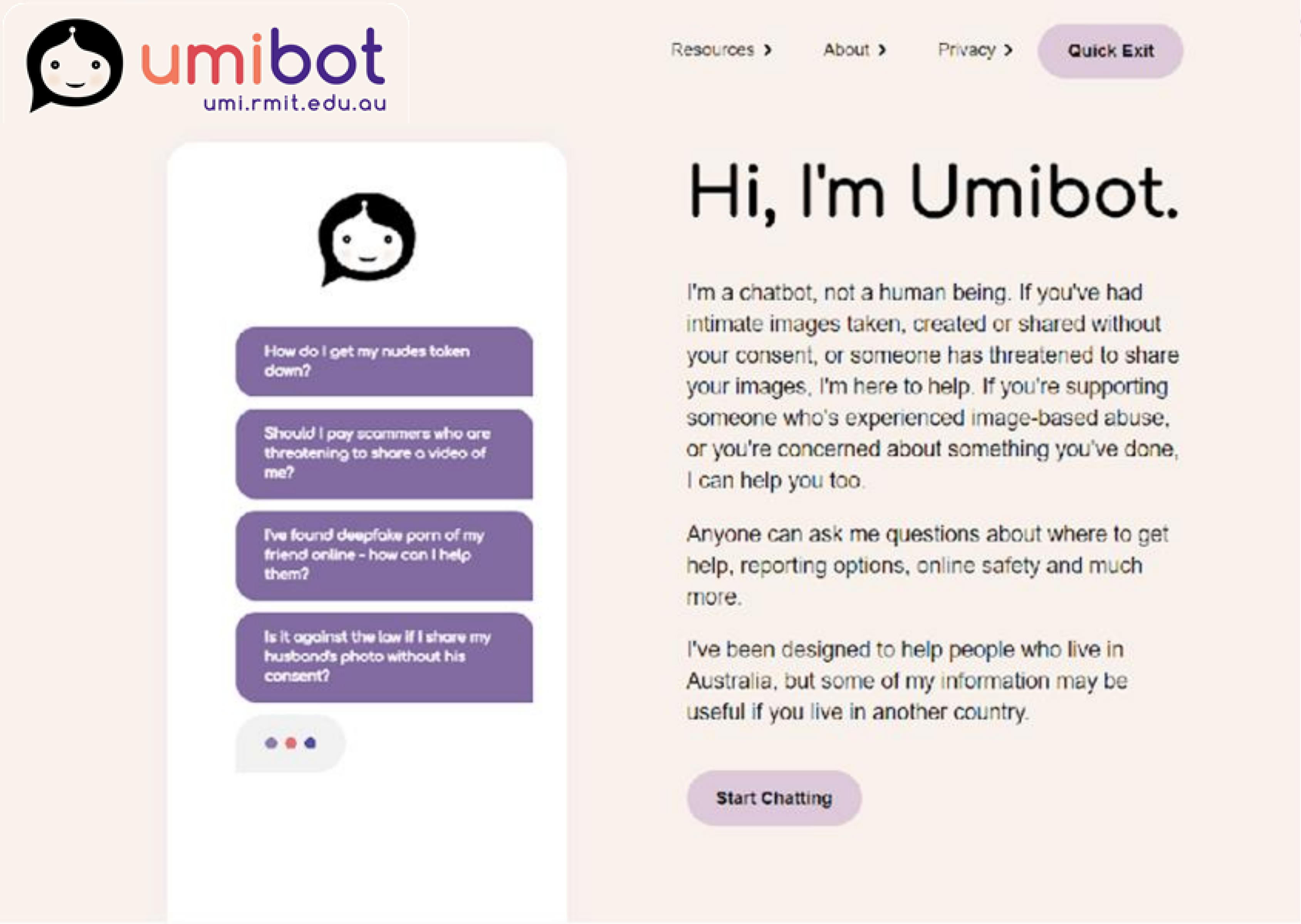
- Consent, gender norms & digital sexual ethics
- Porn / digital sexual literacy
- Challenging 'fun/joke" minimisation narratives
- Targeted education for young people & men

User & Survivor Support

- Reporting pathways (including anonymous reporting)
- Signals & consequences for users
- Empowering users & bystanders
- Access to survivor-centred support



FIGHTING AI WITH AI



How do I get my nudes taken down?

Should I pay scammers who are threatening to share a video of me?

I've found deepfake porn of my friend online - how can I help them?

Is it against the law if I share my husband's photo without his consent?

Hi, I'm Umibot.

I'm a chatbot, not a human being. If you've had intimate images taken, created or shared without your consent, or someone has threatened to share your images, I'm here to help. If you're supporting someone who's experienced image-based abuse, or you're concerned about something you've done, I can help you too.

Anyone can ask me questions about where to get help, reporting options, online safety and much more.

I've been designed to help people who live in Australia, but some of my information may be useful if you live in another country.

- **An informational, text-based chatbot**
- **Aim:** To provide information, support and general advice to victim-survivors, bystanders and, to a lesser degree, perpetrators of IBSA.

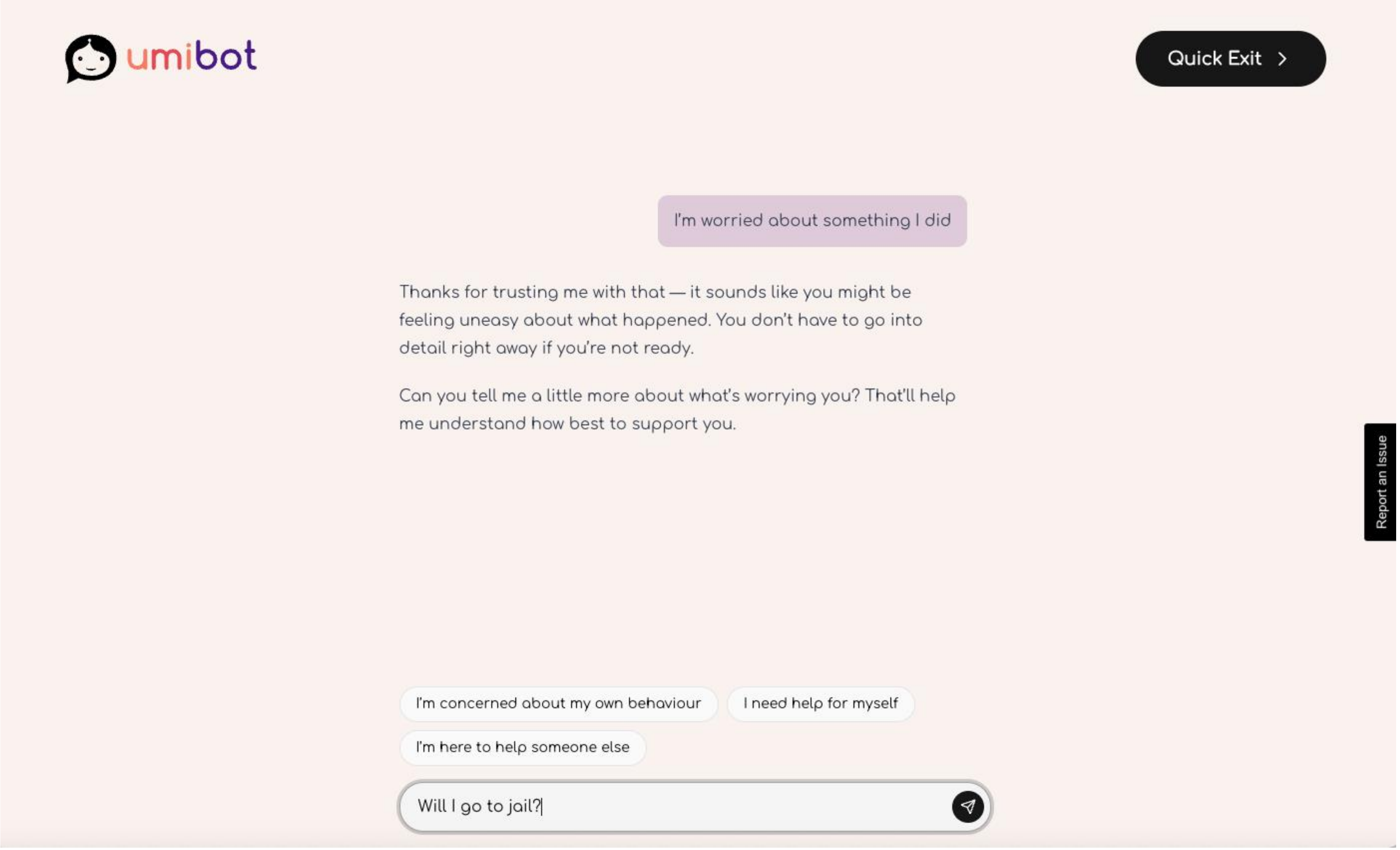
Microsoft Teams

test

2026-02-02 06:09 UTC

Recorded by
Nicola Henry

Organized by
Nicola Henry

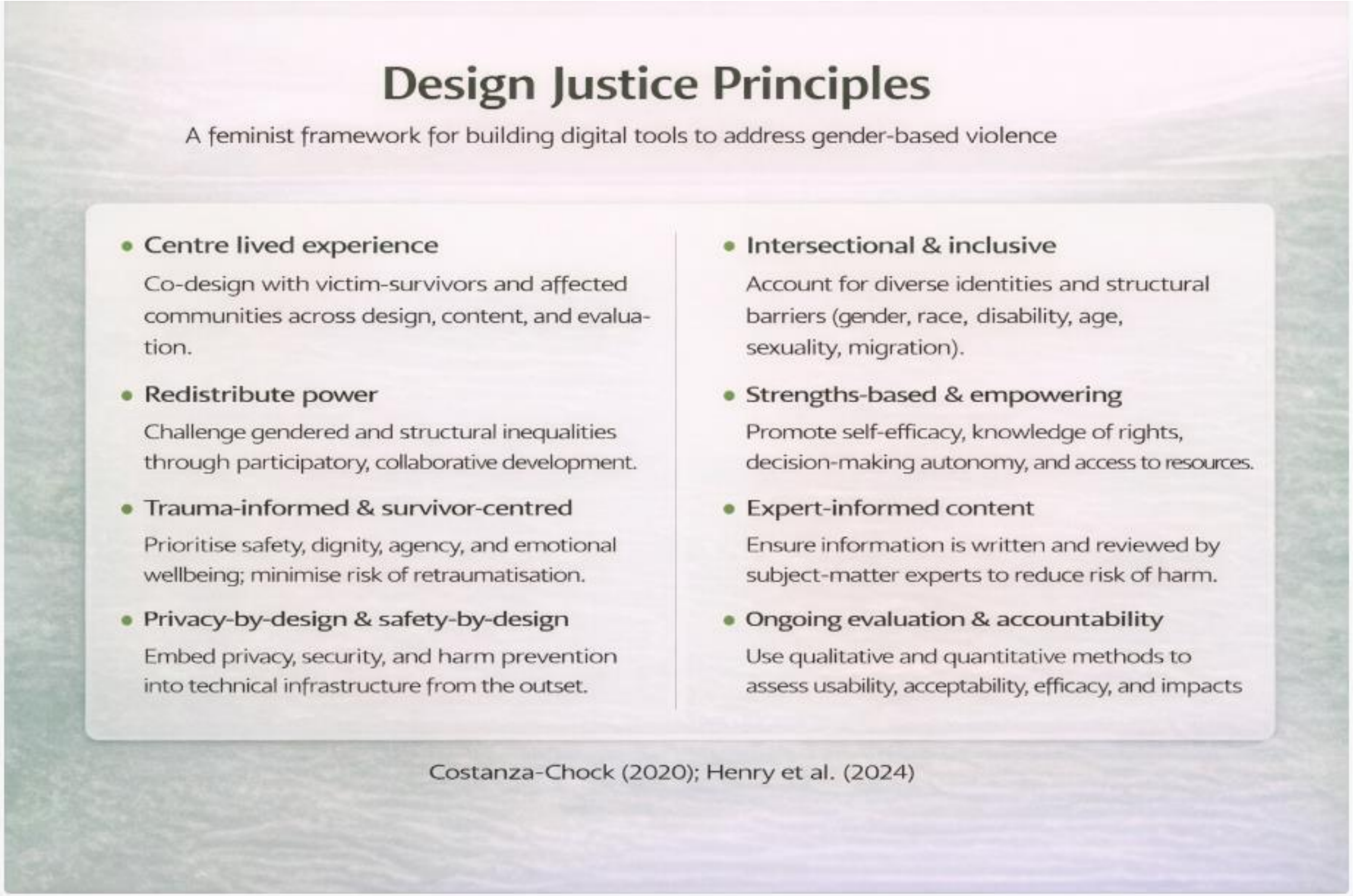

umibot
Quick Exit
Report an Issue
I'm concerned about my own behaviour
I need help for myself
I'm here to help someone else

I'm worried about something I did

Thanks for trusting me with that — it sounds like you might be feeling uneasy about what happened. You don't have to go into detail right away if you're not ready.

Can you tell me a little more about what's worrying you? That'll help me understand how best to support you.

Will I go to jail?



Design Justice Principles

A feminist framework for building digital tools to address gender-based violence

- **Centre lived experience**
  Co-design with victim-survivors and affected communities across design, content, and evaluation.
- **Redistribute power**
  Challenge gendered and structural inequalities through participatory, collaborative development.
- **Trauma-informed & survivor-centred**
  Prioritise safety, dignity, agency, and emotional wellbeing; minimise risk of retraumatisation.
- **Privacy-by-design & safety-by-design**
  Embed privacy, security, and harm prevention into technical infrastructure from the outset.
- **Intersectional & inclusive**
  Account for diverse identities and structural barriers (gender, race, disability, age, sexuality, migration).
- **Strengths-based & empowering**
  Promote self-efficacy, knowledge of rights, decision-making autonomy, and access to resources.
- **Expert-informed content**
  Ensure information is written and reviewed by subject-matter experts to reduce risk of harm.
- **Ongoing evaluation & accountability**
  Use qualitative and quantitative methods to assess usability, acceptability, efficacy, and impacts

Costanza-Chock (2020); Henry et al. (2024)

Conclusion

nicola.henry@rmit.edu.au