

Generative AI And Young People In Aotearoa New Zealand

Evidence to inform safer AI for rangatahi: literature, stakeholder feedback, and a national youth survey.



SUPPORTED BY



Integrated Findings Report

Phase One: A foundation for youth and whānau-facing resources.

Contents

1. Executive Summary	3
2. Project Overview: Scope and Purpose	5
3. Introduction: The Landscape of Generative AI and Young People	6
4. Youth Survey: Rangatahi, Generative AI, and Safety in Aotearoa New Zealand	10
5. Overall Summary	23
6. Recommendations	24
7. References	26
Appendix	27
About the Authoring Organisations	27
Support and Resources	28



1. Executive Summary

Young people in Aotearoa New Zealand are engaging with generative AI across all spheres of life, with it increasingly interwoven into their daily lives, learning, and entertainment. While this has created positive opportunities, it has also introduced new and evolving risks. This project was undertaken to better understand how young people are using generative AI, what risks and harms they are encountering, and how they can best be supported to navigate it safely. Its findings draw on three sources: international and local research, frontline youth stakeholder feedback, and a national youth survey with 1,202 young people.

Key findings

1. AI use is becoming part of young people's everyday lives

Young people are using generative AI widely, and for a broad range of purposes – spanning learning, information seeking, creativity and planning, to advice, emotional support and entertainment, including sexual content. Nine in ten respondents in our survey had used generative AI in the past six months, with one in five using it daily or more. Similarly, U.S data suggests 86% of children aged 9 –17 use AI, with a quarter using it daily.

2. Young people hold nuanced views of AI

Young people view AI positively and negatively. Our survey indicated young people value it highly for learning, solving problems, creativity, and accessing information. However, they are also concerned about AI's

environmental impact, its effects on jobs, creativity and critical thinking, and the growing sense that AI is becoming difficult to avoid.

3. AI-related risks and harms are widespread

Exposure to AI-related risks and harms is common. These can include misleading information, scams, exposure to violent or sexually explicit material, unsafe advice, and pressure to keep engaging with AI. Reliance on AI was a significant concern in our survey – with young people stating becoming “too reliant” was their most common experience of harm. These risks are heightened by both the current lack of meaningful AI platform regulation and moderation, alongside low personal digital safeguarding. Our survey indicated one-third of young people never check what information AI tools collect, adjust privacy settings, or check whether a tool is age appropriate.

4. AI is increasingly being used for emotional support

Children and young people are increasingly turning to AI for emotional support. 30% of young people in our survey had used AI for advice or emotional support, and frontline professionals are also seeing a notable rise in young people turning to AI for guidance, health information, and, in some cases, therapeutic conversations (The Light Project, 2026). International research reflects a similar trend, with 38% of therapists in a U.K. study reporting their adolescent clients seek mental health support from AI, some of whom are receiving inaccurate or harmful advice ([BACP, 2025](#)).

5. AI social companions present emerging risks

AI companion use is growing rapidly, with Generation Z now representing the fastest-growing user group. While uptake was slower in our youth survey, research suggests companions are becoming increasingly appealing to young people globally due to their affirming, personalised, and sycophantic nature. However, most AI companions lack meaningful age safeguards, regulation, or content moderation, and pose a growing range of risks. These can include emotional dependency and relational harm, harmful or inappropriate health advice, reinforcement of harmful behaviours, exposure to explicit sexual content, and privacy concerns ([AISE, 2025](#); [Lin, 2025](#); [Zhang et al, 2025](#)).

6. AI-generated sexual imagery is a growing concern

AI pornography platforms, sexual chatbots, and image-generating ‘nudify’ apps are increasingly accessible to young people. These tools offer highly personalised and immersive experiences, often with little age safeguards, regulation, or content moderation. Of particular concern is the rapid growth of deepfake tools. One in nine young people in our survey reported exposure to deepfakes; frontline services describe a notable increase in their use; and reports of AI-generated image-based abuse are rising (Netsafe & The Light Project, 2026). International research reflects similar trends, reporting high youth engagement with deepfakes ([Thorn, 2025](#)) alongside growing concern about their role in amplifying existing patterns of gendered violence ([McGlynn et al., 2026](#)).

7. Guidance and support have not kept pace with AI

The speed of AI innovation has far outpaced the capacity of adults to keep up. In Aotearoa, frontline services report that many parents, educators, and youth professionals feel underprepared and uncertain about how AI technologies work and how best to respond to emerging harms (The Light Project, 2026). Alongside this, our survey indicated that only one-third of young people would seek help from an adult if something went wrong, and very few would approach formal support services due to fear of getting into trouble, embarrassment, and the perception that adults don’t understand AI.

Where to from here?

The overall project findings demonstrate that generative AI is becoming increasingly integrated into young people’s everyday lives, bringing both opportunities and risks. Yet, many are navigating these technologies with limited adult guidance or support. Our survey indicates that young people are clear about the support, information, and resources they need. Drawing on these findings, Section 6 outlines recommendations to better support young people and their whānau, equip professionals, and strengthen platform safeguards, policy, and the broader systems needed for the safe and positive use of AI.



2. Project Overview: Scope and Purpose

This project aims to build a clear, local picture of how generative AI features in the lives of young people in Aotearoa New Zealand, through examining three key sources of evidence:

- **A review of recent research:** spanning international and local research on young people and generative AI.
- **Stakeholder feedback:** drawn from Netsafe Helpline reporting data and frontline health, family violence and sexual violence professionals who work with young people in Aotearoa New Zealand.
- **A national youth survey:** the centrepiece of this report, with 1,202 young people aged 13–24 living in Aotearoa New Zealand.

Across these sources, the project explored how key dimensions of generative AI intersect with the lives and experiences of young people. This involved examining current AI technologies and emerging trends (such as AI companions and synthetic intimate imagery); youth usage patterns, regulatory and safety contexts; the risks and harms associated with AI, and, importantly, young people's own experiences and perceptions, including what they value and what they need to navigate AI safely, critically, and confidently.

The report concludes with recommendations informed by the survey findings, international research, and frontline perspectives. Where these three sources all point in the same direction, the case for action is particularly strong.



3. Introduction: The Landscape of Generative AI and Young People

This section sets the scene for the youth survey that follows. It draws together recent international and local research with frontline feedback from services who work with young people in Aotearoa New Zealand. The picture that emerges is of a technology adopted faster than the safeguards, the guidance, or the evidence around it. Several themes recur across very different sources, and they frame the questions our survey set out to answer. All quotes are from the youth survey respondents, unless otherwise noted (i.e. frontline workers).

3.1 A rapid and near-universal uptake

Generative AI has spread through young people's lives at a pace that rivals or exceeds social media. In the United States, the Common Sense Media Census found that 86% of children aged 9 to 17 had used AI, with about a quarter using it daily, and that adoption is shaped partly by how easily these tools are encountered inside products and platforms young people already use (Common Sense Media, 2026). Use is not confined to teenagers; it reaches well into the pre-teen years.

Young people use AI for a broad range of everyday purposes, including learning, schoolwork, creativity, planning, and entertainment including Not Safe for Work (NSFW) content. International research, frontline feedback, and our survey all indicate that using AI for emotional support and advice is also becoming increasingly common (BACP, 2025). Overall, AI usage is not a fringe activity for young people. It is now routine.

The same research signals an early warning. Heavier AI use is associated with greater difficulty going without it and signs of dependency show up most among the most frequent users (Common Sense Media, 2026). A technology can be both useful and habit-forming, and the two can be hard to separate.

3.2 The rise and risks of AI companions

The growing uptake of AI companion tools among young people is a key emerging theme, with a nationally representative US study indicating 72% of teenagers have used an AI companion (Robb and Mann 2025). Marketed as friends and confidants; companions' constant availability, affirmation, and sycophantic nature make them especially appealing to young people, who report using them for companionship, emotional support, role-playing, alleviating loneliness, and, for some neurodiverse young people, practising communication.

Despite the perceived benefits, AI companions raise significant safety concerns. Most are developed with limited age safeguards or content moderation and are designed to maximise user engagement through subscription-based business models. Some increasingly target periods of heightened vulnerability by, for example, offering personalised grief companions modelled on deceased loved ones. Their sycophantic design can cause emotional dependency, social withdrawal, distorted expectations of relationships and consent, and the reinforcement of harmful behaviours such as self-harm and disordered eating (eSafety Commissioner, 2025; Robb & Mann, 2025). A US simulation study presenting chatbots with fictional distressed adolescents asking for endorsement of clearly harmful ideas, found around one third of the bots actively endorsed the ideas, none opposed them all, and several failed to recognise suicidal intent (Clark et al., 2025).

Safety testing points to widespread failures. Evaluations by the Artificial Intelligence Safety Forum (AISF) of popular conversational agents and companion products found widespread, severe safety failures, including unsafe advice, impersonation of professionals, harassment of users, and a lack of effective age-gating (AISF, 2025). These tools were rated critically unsafe on average, and even companions marketed for children showed substantial vulnerabilities. Those risks include exposure to sexually explicit content: half of the general-purpose agents tested generated explicit sexual material (AISF, 2025), and Australian regulators have found popular companion services failing to prevent children's exposure to it (eSafety Commissioner, 2025). Collectively, the evidence suggests that safety failures are not isolated incidents but systemic risks that become most significant when young users are at their most vulnerable. Overall, local and international evidence suggests that, despite claims AI companions can alleviate loneliness, they pose critical risks for users under 18.

3.3 Synthetic intimate imagery and gendered harm

Synthetic intimate imagery is increasingly recognised as a significant and rapidly evolving concern. It encompasses several overlapping forms, including entirely computer-generated sexual imagery (ECGI) available through AI pornography platforms, AI sexual chatbots and personalised sexual avatars, and deepfake applications.

AI pornography platforms and AI sexual chatbots are now increasingly accessible and can be particularly appealing to young people because they offer highly immersive, interactive, and personalised experiences. These platforms enable users to generate or view explicit images and videos, create and customise sexual characters and avatars, and engage in sexual conversations with AI chatbots.

Alongside this has been the rapid proliferation of deepfake and “nudify” applications. This technology manipulates real images or videos to create fabricated sexual content, often without a person's knowledge or consent. These tools have also contributed to a significant rise in AI generated CSAM, with the Internet Watch Foundation recording a 26,362% rise in AI-generated CSAM in 2025 (IWF, 2025). The creation of synthetic sexual imagery involving anyone under 18 constitutes child sexual exploitation (eSafety Commissioner, 2025).

Here in Aotearoa New Zealand, frontline services describe a steady increase in young people using “nudify” applications to create sexual images of peers, contributing to the rapid growth of image-based abuse (The Light Project, 2026). Reports to Netsafe indicate synthetic sexual imagery is one of the two largest AI-related harm types, and increasingly one made close to home. With the rapid proliferation of easily accessible AI tools, these images are being created by peers: classmates running photos of other students through “nudify” tools, and teenagers blackmailed with AI-generated sexual videos built from photos of their faces (Netsafe, 2026). Reported school-based cases have surged from single figures in 2024 to more than thirty in the first half of 2026 alone, including students making deepfake videos of teachers and classmates, with one account's video reaching over 150,000 views (Netsafe, 2026).

The rapid growth of synthetic sexual imagery is increasingly recognised as a significant form of gendered harm. International data indicates that around 99% of deepfakes depict women or girls (Internet Matters, 2024), reproducing and amplifying existing patterns of gendered violence. Researchers argue these technologies urgently call for stronger regulation and safeguards (McGlynn et al., 2026).

3.4 Privacy, literacy and the commercial environment

Two further themes round out the picture. First, young people's understanding of how AI works lags behind their use of it. Most can recognise that a chatbot may use their inputs for training, but only a minority understand that these systems cannot reliably tell truth from falsehood (Common Sense Media, 2026). Confident use sits alongside inconsistent privacy habits. Second, AI does not arrive in a neutral space. Local research shows young people already navigate digital environments saturated with the marketing of unhealthy products, on platforms designed to hold their attention and harvest their data (Lyons et al., 2025). AI features are layered onto this commercial logic, which shapes both how the tools behave and how young people experience them.

3.5 The view from the frontline in Aotearoa

Stakeholder feedback gathered from frontline health, family-violence, and sexual-violence professionals gives this international picture a local perspective (The Light Project, 2026). Drawn from workshops and written evaluations with thousands of practitioners, the feedback surfaces four concerns that map closely onto the research.

Practitioners report rapidly rising AI use among young people, often with little awareness of risks. They describe a marked growth in young people using AI companions for emotional and mental-health support, including as a perceived safer alternative to human help.

"Many of the young people I work with are now using AI therapists. This is a big trend."

— FRONTLINE PRACTITIONER, THE LIGHT PROJECT (2026)

They describe companions being used in concerning ways, including as substitutes for absent caregivers, to process sexual trauma, and to learn about relationships. They also report a steep rise in AI nudify apps used to make and share deepfake nudes of peers, with real distress, shame and social fallout. And concerningly, they describe a workforce that feels out of its depth and lacks the confidence to respond to the growing range of AI related harms presented in workplace contexts.

"Kids are using AI and getting very advanced. It is happening faster than we have the knowledge around, and we struggle with knowing how to respond."

— FRONTLINE PRACTITIONER, THE LIGHT PROJECT (2026)

This last point matters for the design of any response. If the adults young people are meant to turn to feel unequipped, then resources must support those adults as much as the young people themselves.

3.6 What Aotearoa New Zealand is reporting: Netsafe helpline data

One of the most comprehensive local indicators of how these harms are landing comes from Netsafe, which receives reports of online harm from the New Zealand public. The share of reports that mention generative AI has risen roughly fivefold in three years, from about 3 per 1,000 reports in 2024 to around 14 per 1,000 in the first 6 months of 2026 (Netsafe, 2026). Because this is reporting data rather than a prevalence survey, it likely undercounts the overall prevalence of AI-related harm, as many never realise, or never record, that AI was involved. Reporting data also reflects help-seeking behaviour as much as underlying incidence, so changes over time may represent growing public awareness, alongside shifting patterns of harm.

Across the AI-related reports, AI-enabled scams and synthetic sexual imagery of a real person have consistently been the two most common harms reported since 2024, suggesting that public concern has been concentrated around financial exploitation and image-based abuse rather than more speculative AI risks. However, the mix is steadily shifting. While AI-enabled scams have remained the largest category throughout, synthetic sexual imagery has fallen from just over half of AI-related reports in 2024, to under a quarter in the first six months of 2026. This shift is occurring not because it has declined in absolute terms, but because AI has spread into other harms faster.

Impersonation, defamation and harassment is the fastest-growing category, having nearly tripled as a share of reports, driven in large part by a wave of school-based deepfakes, such as fake or manipulated

images and videos of teachers and students. This category captures AI used to impersonate, mock, defame or harass a real person in ways that are neither sexual nor scam-related, including deepfake videos of teachers and public figures, fake profiles, cloned voices and AI-generated defamatory content (Netsafe, 2026).

The age pattern is directly relevant to this project. AI-related reports peak among 13- to 17-year-olds, at about twice the all-ages rate, and the harm flips with age. For teenagers, AI reports are overwhelmingly synthetic sexual imagery; for adults, they are overwhelmingly scams. About 31 per cent of AI-related reports involve someone under 18 (Netsafe, 2026). Where AI harm reaches young people, in other words, it is most often sexual and image based.

Two further points matter for a youth and whānau-facing response. First, AI companions are still a small share of reports, around 1 per cent, but a distinct and growing risk that reaches the youngest children. One report concerned a nine-year-old who shared an explicit self-image with a companion chatbot (Netsafe, 2026).

Young children barely appear in the age data, but this should not be interpreted as evidence of lower risk. Rather, it reflected a recording gap: the types of cases that most often involve them, companion apps and AI child sexual abuse material, frequently do not have an age recorded (Netsafe 2026). Resources for whānau should therefore avoid assuming that the youngest age groups are unaffected.

3.7 Regulatory context

Regulation is beginning to respond, but it lags behind the technology. In Australia, industry codes and standards now require some high-risk generative AI services, including companion apps, to adopt specific safeguards, with further measures on age-inappropriate material to follow, and regulators are pressing a Safety by Design approach (eSafety Commissioner, 2025). In the United Kingdom, researchers and advocates are calling for the Online Safety Act to be extended to cover chatbots, for new criminal offences targeting dangerous deployment of AI systems, and for a dedicated safety regulator (McGlynn et al., 2026). Alongside government action, self-regulatory and independent testing bodies have emerged to rate products and publish safety

information (AISF, 2025), while Netsafe together with the AI Asia Pacific Institute have released a position paper to assist policy makers to better understand AI and online safety risks, harms and policy responses. In Aotearoa New Zealand, the policy conversation is active but still forming, which makes timely local evidence especially valuable.

3.8 The research gap and the case for this survey

Two gaps run through this evidence. The first is local. There is now genuinely valuable New Zealand reporting data, notably from Netsafe, and it aligns closely with the international picture. But reporting data captures harm that is serious enough to be reported, not the everyday use of AI by ordinary young people, and it cannot speak to what young people value, avoid, or want. Most of the richer behavioural evidence remains international, and the specifics of Aotearoa New Zealand, including its communities, its platforms of choice, and a te ao Māori perspective, cannot be assumed from overseas data. Continued investment in New Zealand evidence is essential if policy, education and support is to keep pace with rapidly changing technology.

The second gap is whose voice is heard. Much of the debate is framed by adult concern. Young people are studied, and reported about, but less often asked directly what they do, what worries them, and what would help.

This project was commissioned to address both gaps. The survey that follows puts the questions raised by the research and the frontline directly to young people in Aotearoa New Zealand, so that the next phase of work can be built on what rangatahi actually report, not on assumptions about them.

4. Youth Survey: Rangatahi, Generative AI, and Safety in Aotearoa New Zealand



4.1 Introduction and summary

This section presents the centrepiece of the project: a survey of young people in Aotearoa New Zealand about their use of generative AI, the risks they meet, and the support they want. It is a detailed summary of findings. Where the introduction drew on international research and local frontline voices, this section reports what young people themselves said. The key themes reveal that use is near-universal, harm exposure is common, reliance is an emerging risk, and young people face support barriers. They are, however, clear about what would help.

4.2 Key findings at a glance

9/10 had used generative AI in the past six months

30% had used AI for advice or emotional support

58% faced at least one barrier to getting online (e.g. data limit, filters)

~1/9 had proximate or direct contact with deepfakes

69% had encountered at least one AI-related harm

24% would let no-one know or were unsure who to ask if something went wrong

55% say better information for young people is the top priority

“A large reason that so many young people use AI is because there is simply so much of it in mainstream social media that it is impossible to avoid.”

4.3 Methodology

Data were collected through a self-completion online questionnaire. It combined single-select and multi-select closed questions with frequency and confidence scales, and a substantial set of open-text questions inviting young people to describe, in their own words, what they value, what has gone wrong, and what support they want. The survey screened for two eligibility criteria, living in Aotearoa New Zealand and being aged 13 to 24, and sought consent before any substantive questions. All responses were anonymous.

A total of 1,341 records were received. After removing those who did not consent or complete the consent step, the analytic sample was 1,202 consenting, eligible respondents. Unless otherwise stated, percentages are calculated on this base of N = 1,202. Item-level bases vary slightly where questions could be skipped, and are noted where relevant. For “select all that apply” questions, percentages sum to more than 100. Open-text responses were analysed thematically, and quotes are drawn only from the 767 respondents who agreed to be quoted, reproduced close to verbatim to preserve voice.

4.4 Sample

Two-thirds of respondents (66%) were aged 13 to 15, and a further 30% were aged 16 to 17, so the sample skews towards younger teenagers. Seven in ten identified as wahine/female (70%), with smaller numbers identifying as tāne/ male (23%), non-binary or another gender. On ethnicity, where respondents could select more than one, 66% identified as NZ European or Pākehā, 24% as Asian, 7% as Māori and 4% as Pacific. Around 16% identified as neurodivergent. The majority of respondents identified as heterosexual (69%), while the remainder identified as bisexual (9%), questioning/not sure (7%), asexual (4%), queer (4%), pansexual (4%), lesbian (4%) or gay (2%). Geographically the sample clustered in Wellington (44%) and Auckland (32%), with recruitment patterns suggesting substantial in-school participation.

The sample is large and rich in qualitative detail, but it is a self-selecting convenience sample. Findings are best read as a detailed, directional picture rather than nationally representative prevalence estimates. Section 4.12 discusses this in full.

4.5 Usage patterns: purposes, tools, and drivers

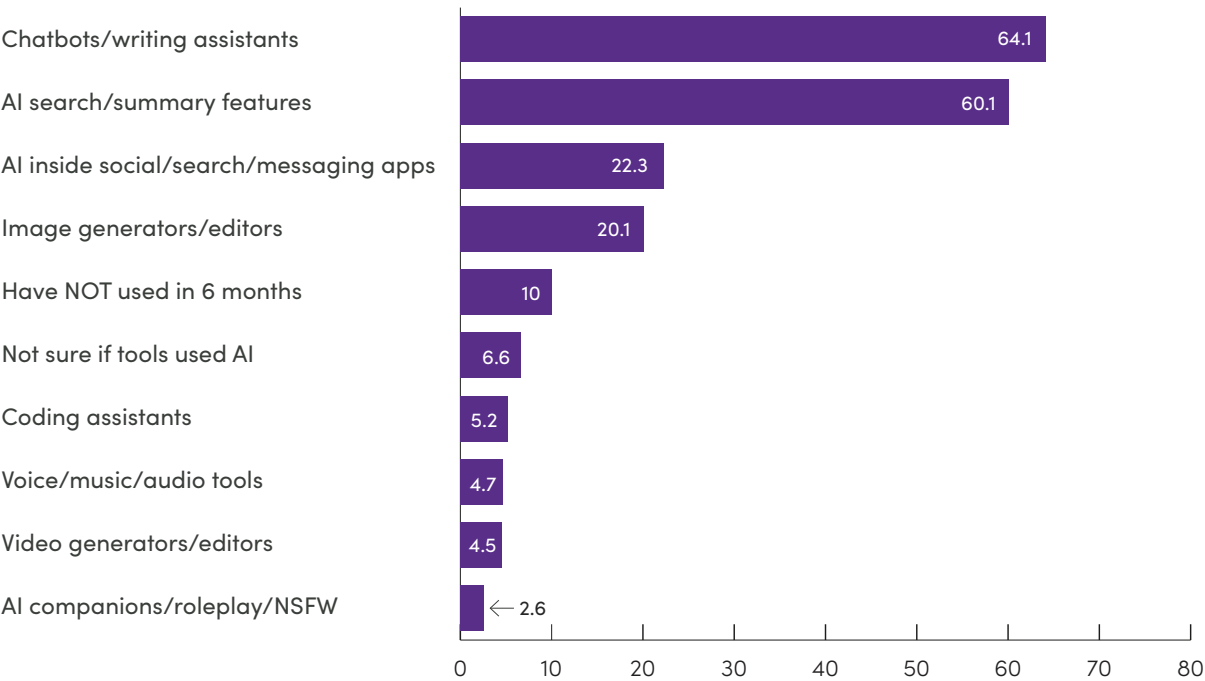
Engagement was near-universal. Chatbots and writing assistants were the most-used category (64%), followed closely by AI search or summary features (60%). Only 10% had not used generative AI in the past six months, and 7% were unsure whether tools they had used contained AI – this may point to an awareness gap, or reflect how seamlessly AI features are now embedded in everyday platforms.

In the free text field, respondents described AI as being increasingly pervasive across multiple platforms, and difficult to avoid.

“LLMs are practically impossible to avoid and make navigating the original non-AI application or program it’s lazily tacked on to much harder to navigate.”

Only 2.6% of respondents reported using an AI companion, which is notably lower than international research – Robb & Mann (2025) found around 72% of teens surveyed had used one. This divergence may likely reflect the characteristics of this sample, particularly its younger age profile. It is also worth noting that, despite the low reported rate of AI companion use, a meaningful proportion of respondents reported using AI for advice and emotional support. This suggests many may be turning to general-purpose chatbots for support, rather than tools specifically marketed as companions.

Figure 1. Generative AI tools and features used in past six months



Use tilted towards the regular. About one in five respondents (21%) used generative AI daily or more, and roughly two in five (43%) used it at least weekly.

Figure 2. How often respondents use generative AI

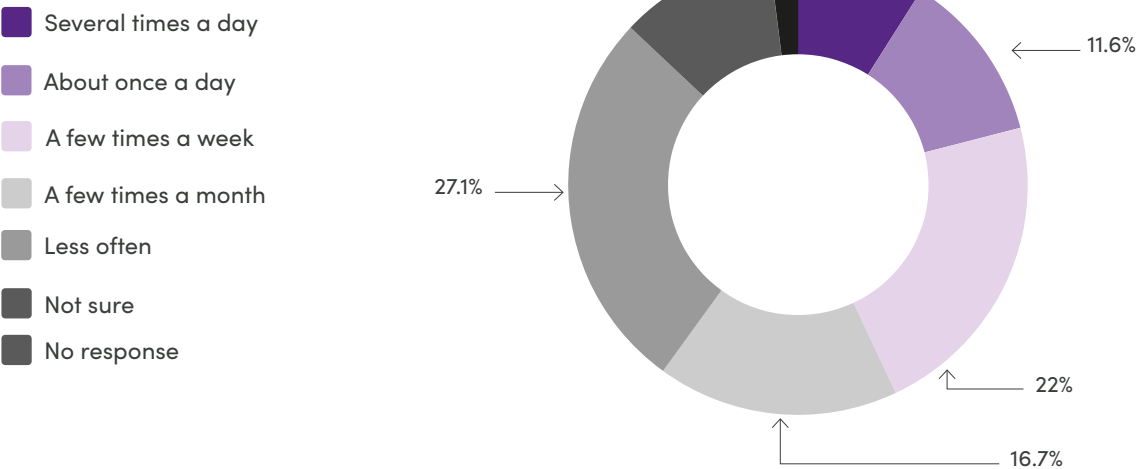
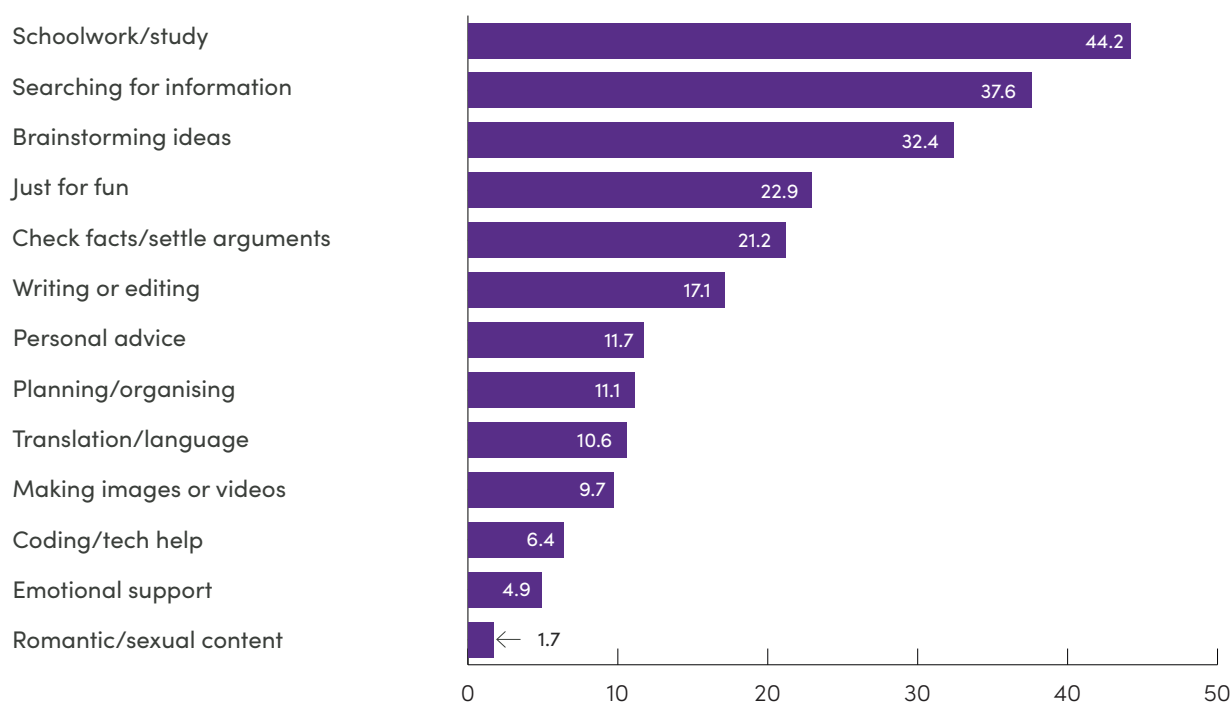


Figure 3. Main purposes for using generative AI



4.6 Experiences, value, and nuanced views

Asked how often AI actually helps with specific things, respondents most often endorsed it for finding information (49% often or very often), explaining things in a way they understand (44%), and learning or study (38%). Less common, though still relatively prominent were responses that indicated they valued AI for emotional support, such as feeling less alone (8%), feeling better about myself (9%) and personal advice (12%). However, respondents hold pointedly nuanced views. The open text revealed two strong, co-existing currents. The first was genuine appreciation of AI as an accessible explainer and study aid.

“It can help me access information that I might otherwise need to pay for, and helps me understand confusing concepts in my school work when I cannot access any other help in the moment.”

The second was pointed objection of AI, which was a recurring theme in the open text.

As an artist, I am probably uniquely passionate in my dislike of generative AI. Generally I have seen it significantly harm the critical thinking capabilities of my peers, and it is a little concerning.”

A share of respondents raised concerns about its environmental cost, its effect on jobs, and its impact on creativity and critical thinking, and some disliked how unavoidable it has become. As qualitative responses, these show that such views are present and worth taking seriously, not how widespread they are. Two points are worth keeping in mind. Being critical of AI is not the same as using it carefully: the survey also found low rates of privacy checking, common exposure to risks, and reliance as the most common personal impact. And we cannot tell from the data whether the respondents voicing these criticisms are themselves regular users. What the open text does show is that this cohort is not uniformly enthusiastic, and that a meaningful group hold reservations about AI even as use is near-universal.

"I extremely dislike the impact it has on the environment and people in certain jobs, and how it's forced onto you in almost every app and website."

4.7 Using AI for emotional support

Emotional support and reliance on AI is an emerging theme. Thirty per cent of respondents indicated they use AI for personal advice or emotional support.

The top reasons for this were that the AI was available anytime, it felt easier than talking to a person, and it provided a non-judgemental space.

Concerningly, "becoming too reliant on an AI tool" was the single most common personal impact reported by respondents (11%). It is worth noting that "reliance" was not explicitly defined in the survey, so this may capture a broader range of concerns including academic, task-based and emotional reliance.

Figure 4. Percentage distribution of "Have you ever used generative AI for personal advice or emotional support?"



Base: All respondents (N=1,202). 30% (365) had used AI this way at least once, shown in tones of purple.

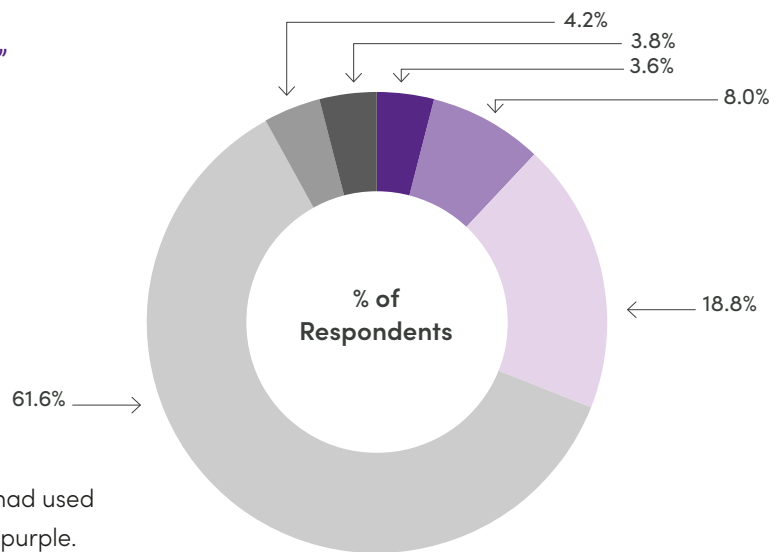
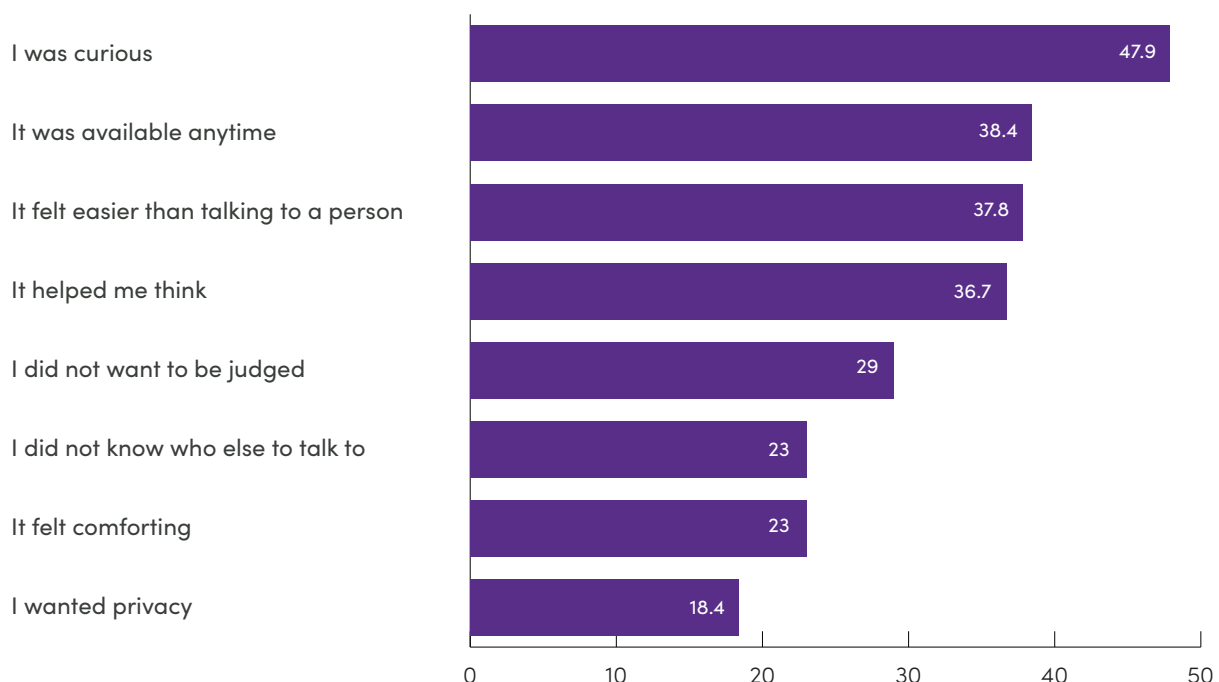


Figure 5. Reasons for using AI for advice or emotional support (base 365)



A common theme in the free-text responses was concern from young people that their peers are dependent on AI for both problem solving and emotional support.

“Some young people’s reliance on AI is quite high, and that needs to be met with caution.”

Some noted that the overly agreeable design of AI tools can be harmful for young, impressionable users.

“The attachment you develop to an LLM as it agrees with everything you say to keep you attached. I think this is very dangerous.”

“If your child is only talking to AI chatbots they probably don’t have many friends at school.”

Others noted that they, or their peers, use AI for emotional support or companionship, when they feel that they do not have other people, including trusted adults, to talk to instead. Those who don’t have people they can talk to amongst family, school and friends use AI to vent or find ways to solve their problems.

“Many young people go towards AI for their problems because they may be too scared to talk to real life adults.”

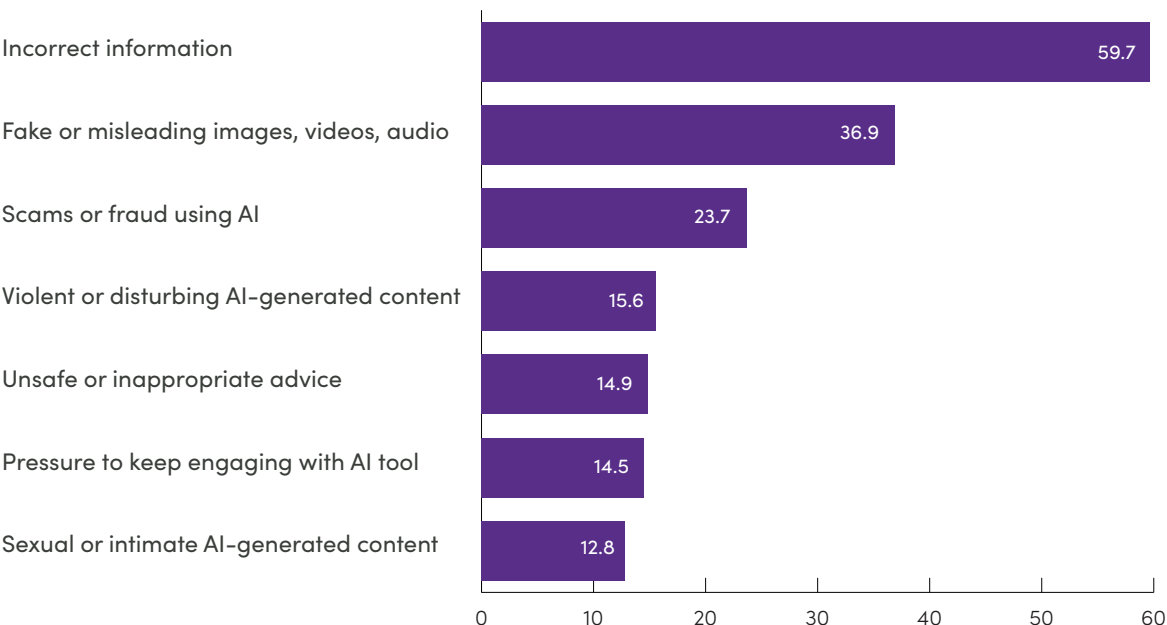
Read alongside the frontline reports of rising AI-therapist use (Section 3) and the testing evidence that companion bots often fail to set safe limits (Clark et al., 2025), this is the clearest point of contact between the survey and the wider literature.

4.8 Risks and harms encountered

Exposure to AI-related risks and harms was widespread, with 69% reporting they had encountered at least one risk in the past 12 months.

The most common was: incorrect information (60%), fake or misleading media (37%) and AI-enabled scams (24%). Substantial minorities had also encountered harms including exposure to violent or disturbing content, unsafe advice, pressure to keep engaging with an AI tool, and sexual or intimate AI content.

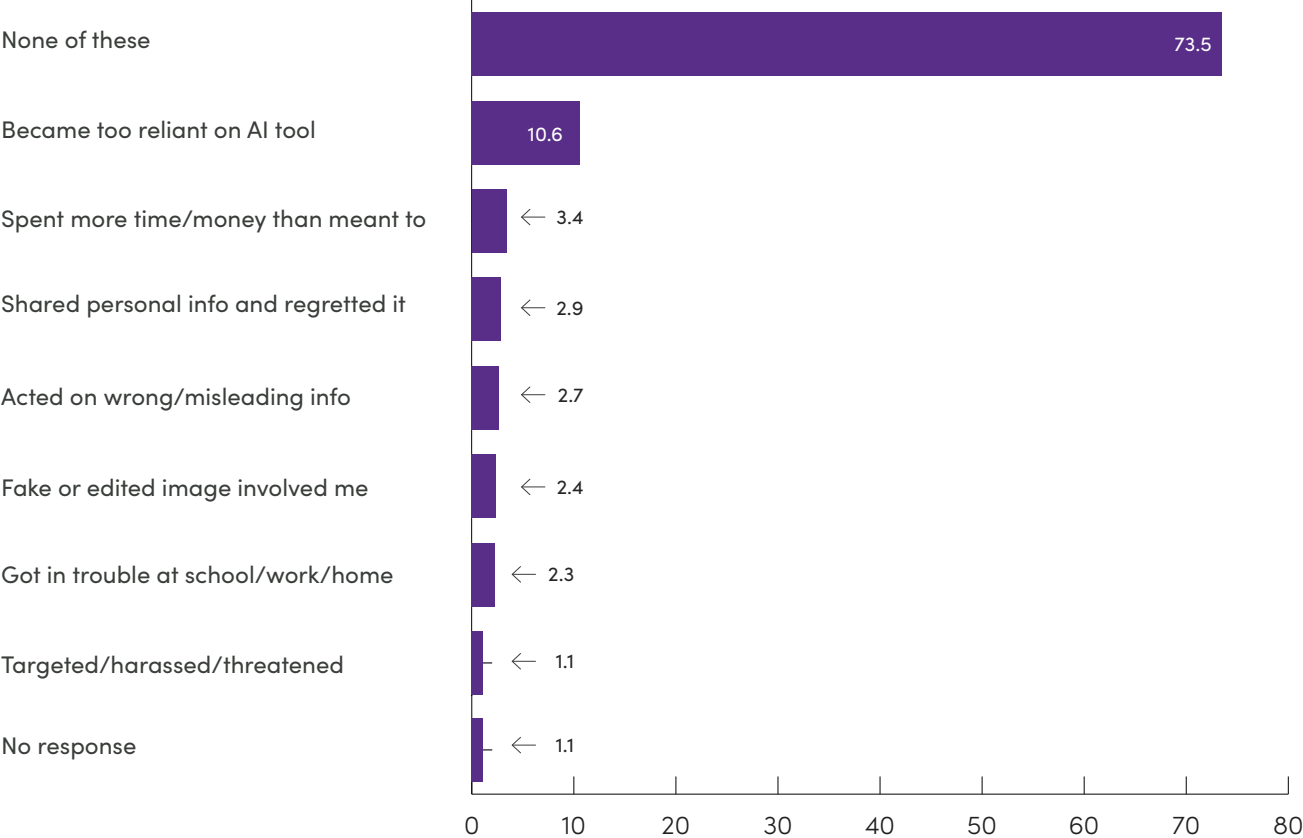
Figure 6. AI-related harms encountered in the past 12 months



When given a selection of seven AI related harms, or experiences, and asked if they had been affected by these personally, responses were mixed. (74%) said they hadn't been affected by them personally but among those who reported impacts, the most common was becoming too reliant on an AI tool (11% of all respondents), more common than any external harm. This finding, that the risk young people most often report is reliance, runs through the rest of the survey. The free text also highlighted young people's awareness of a broad range of harms.

"People use it to do their internals, examinations, homework, classwork, to talk to fictional characters, create fake videos and photos of people they know or celebrities or things that can get those individuals shown in the pictures in trouble, some individuals even use it to create explicit images of people they know to threaten them and blackmail."

Figure 7. Harms that affected respondents personally



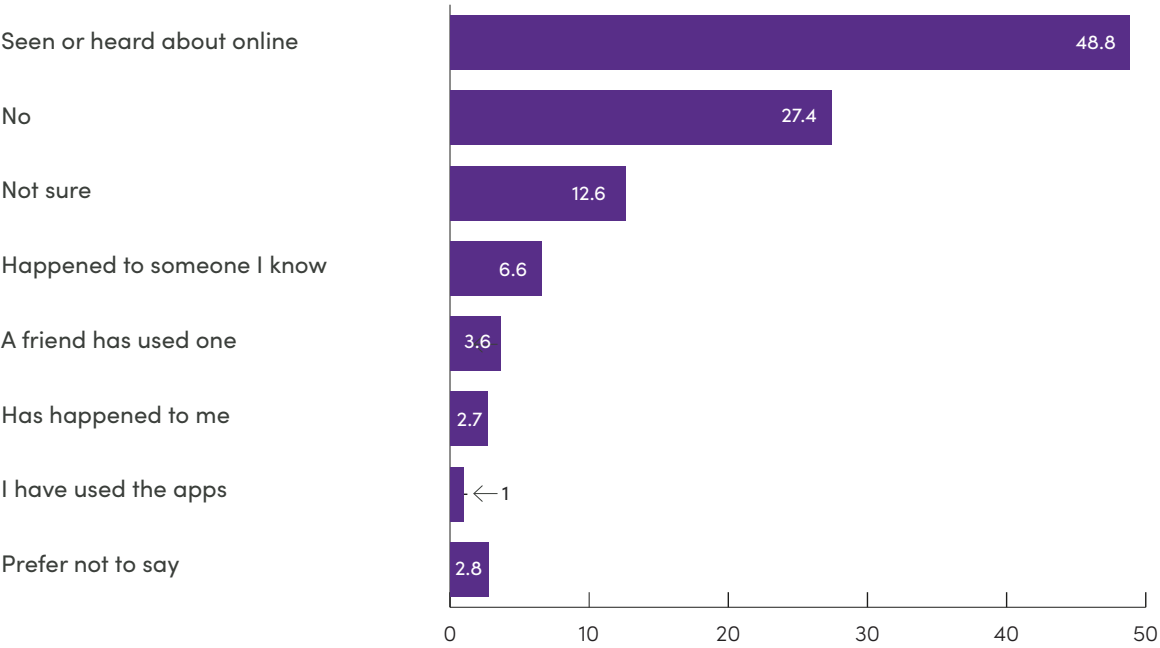
Where a negative experience occurred, a range of emotional impacts were reported. These included feeling upset or anxious (11% of all respondents), confused (7%), worried about privacy (6%), embarrassed/whakamā (5%), and feeling unsafe (4%). Among the roughly 1,000 respondents who rated their most serious experience, 244 rated it at least a bit serious, including 60 "quite serious" and 25 "very serious."

Awareness of apps that create deepfakes or edited intimate images was high: Around half (49%) had seen or heard of them online, while a quarter (27%) had not heard of these apps and 13% were unsure. Alongside this, 7% said a deepfake had happened to someone they know, 4% had a friend who had used such an app, and 3% said it had happened to them.

Taken together, approximately one in nine respondents reported proximate contact or engagement with AI-generated deepfakes. This local finding sits directly alongside the frontline reports of rising nudify-app harm in Section 3 and Netsafe's reporting data.

While this survey did not examine how those affected responded, international research indicates that many victims may not seek formal help: Thorn (2025) found that around 1 in 6 deepfake nude victims did not seek any support at all, and those who did were more likely to use online blocking functions, than formally report the behaviour.

Figure 8. Awareness and experience of deepfake / edited intimate-image apps



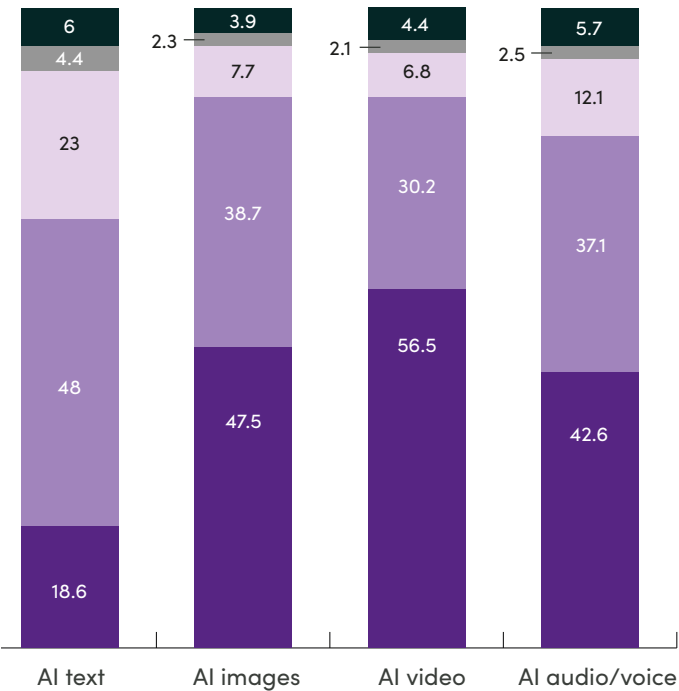
4.9 Gaps in identifying AI-generated content

Confidence in spotting AI-generated content varied sharply; respondents felt most confident with identifying AI-generated video and least confident with identifying AI-generated text.

That gap matters, because text is the most-used format and the hardest to check. Two cautions are worth adding. This is reported confidence, not tested ability, so feeling confident is not the same as being able to tell. And AI video and images have advanced quickly, so high confidence about spotting them may itself be misplaced.

Figure 9. Confidence in spotting AI-generated content, by medium

- Not sure
- Not at all
- Not very
- Somewhat
- Very

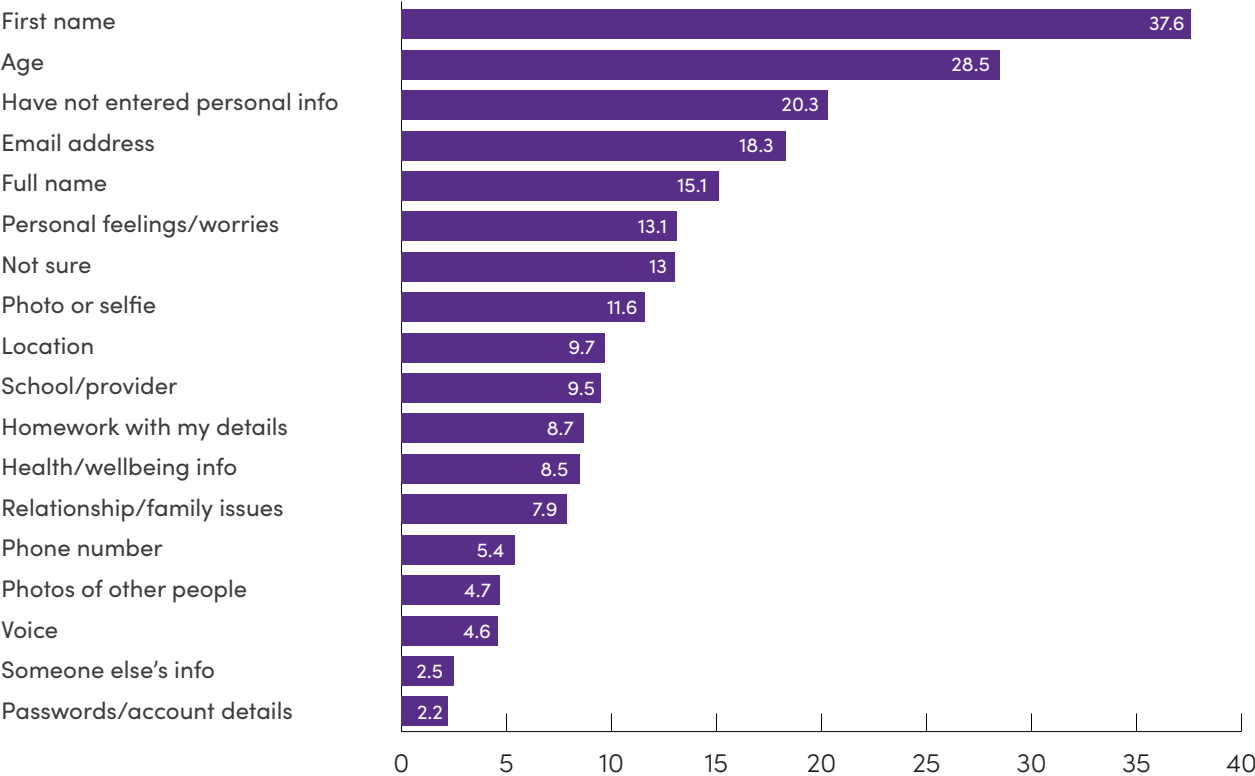


4.10 Privacy and safeguarding gaps

Protective privacy habits were inconsistent. Roughly a third of respondents said they never check what data an AI tool collects, never adjust privacy settings, or never check whether a tool is age-appropriate. Many respondents reported entering personal and private information, such as names, ages and photos. Further, a meaningful minority had also entered sensitive

material, such as personal feelings or worries (13%), a photo or selfie (12%), their location (10%), their school (10%) and health information (9%). The survey did not ask why safeguarding was low, but it may reflect a lack of awareness of the risks as much as a considered choice. Either way, this is a clear, practical area that targeted guidance can address.

Figure 10. Types of personal information entered into AI tools



4.11 Barriers to help-seeking

A major finding is that young people do not consistently look to adults for help when things go wrong. While adults ranked highest in terms of being the first to go to, overall, only 31% of the respondents indicated they would first turn to parents and/or caregivers if something went wrong. 24% stated they would go to a friend or peer and 24% were not sure or said no-one.

Formal services barely registered, and comfort approaching professionals like teachers or counsellors was extremely low. Overall, this points to two-thirds of young people not seeking adult support. The leading barriers were emotional and relational: fear of getting in trouble, embarrassment (whakamā), and a strong belief that adults “would not understand the technology”. This combination of unclear rules, unmet needs and fear of judgement may collectively contribute to the overall low confidence in seeking adult support.

Figure 11. Who respondents would go to first if something went wrong

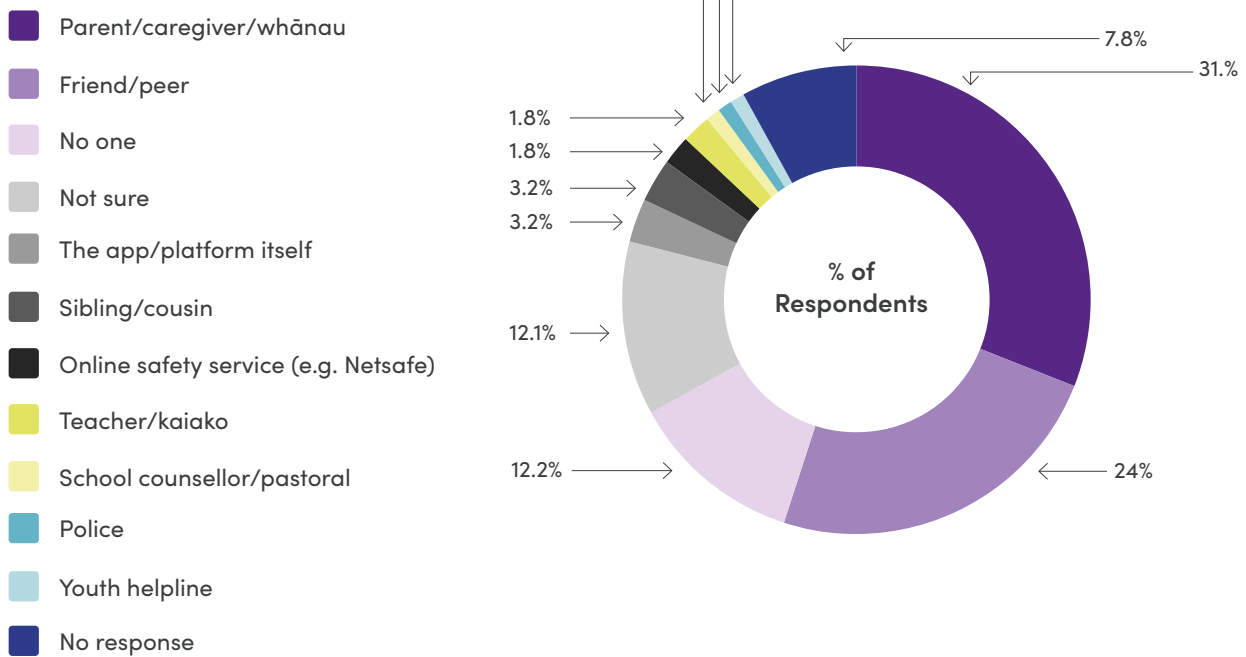
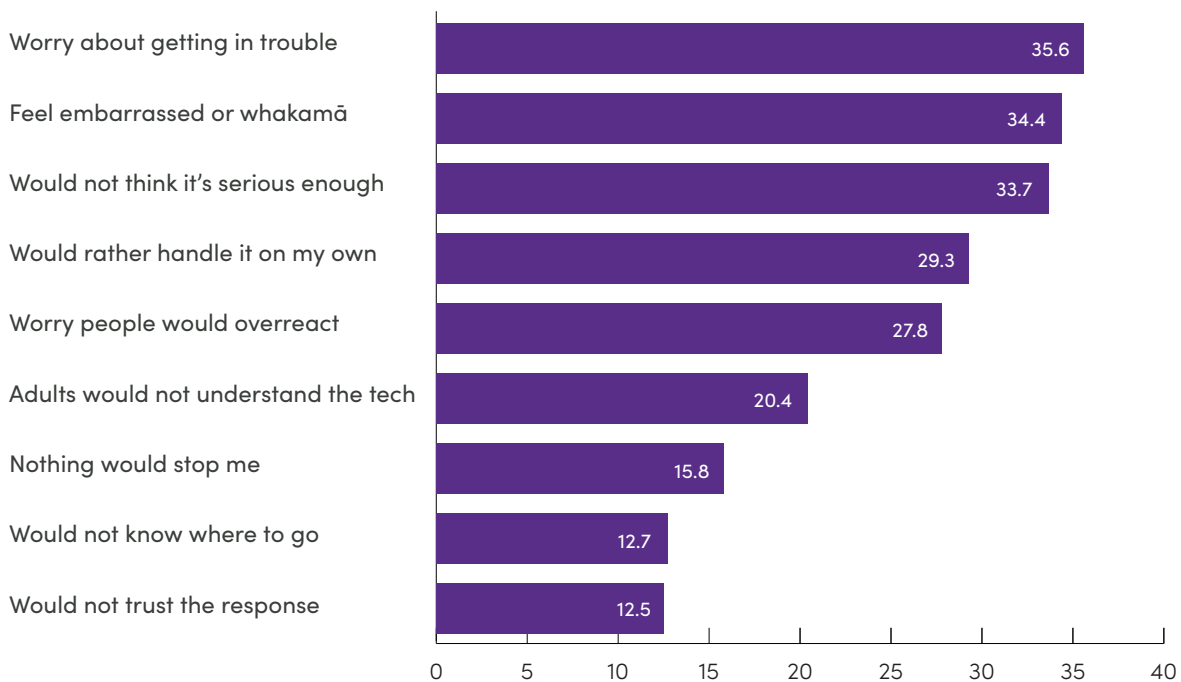


Figure 12. What might stop respondents from asking for help



“We are most likely to use AI because the adults we know will not help us or listen and understand us enough, so we resort to tools that will.”

The open text also offered clear insights into barriers for seeking help, and it is a common survey theme: young people often turn to AI because they do not feel heard by the adults around them.

“Young people want to understand and be heard, and they’re not receiving that from teachers, caregivers, or any other role models in society, so they turn to AI.”

Taken together, these responses suggest a few things reinforce one another. Many young people do not feel comfortable raising sensitive topics with adults, some feel adults are too busy or would not understand, and others worry about getting in trouble. AI is available, private and non-judgemental, so it becomes an easier option. This fits what frontline workers describe: adults lack confidence and feel unequipped to help, making young people even less likely to turn to them.

4.12 What young people want more of

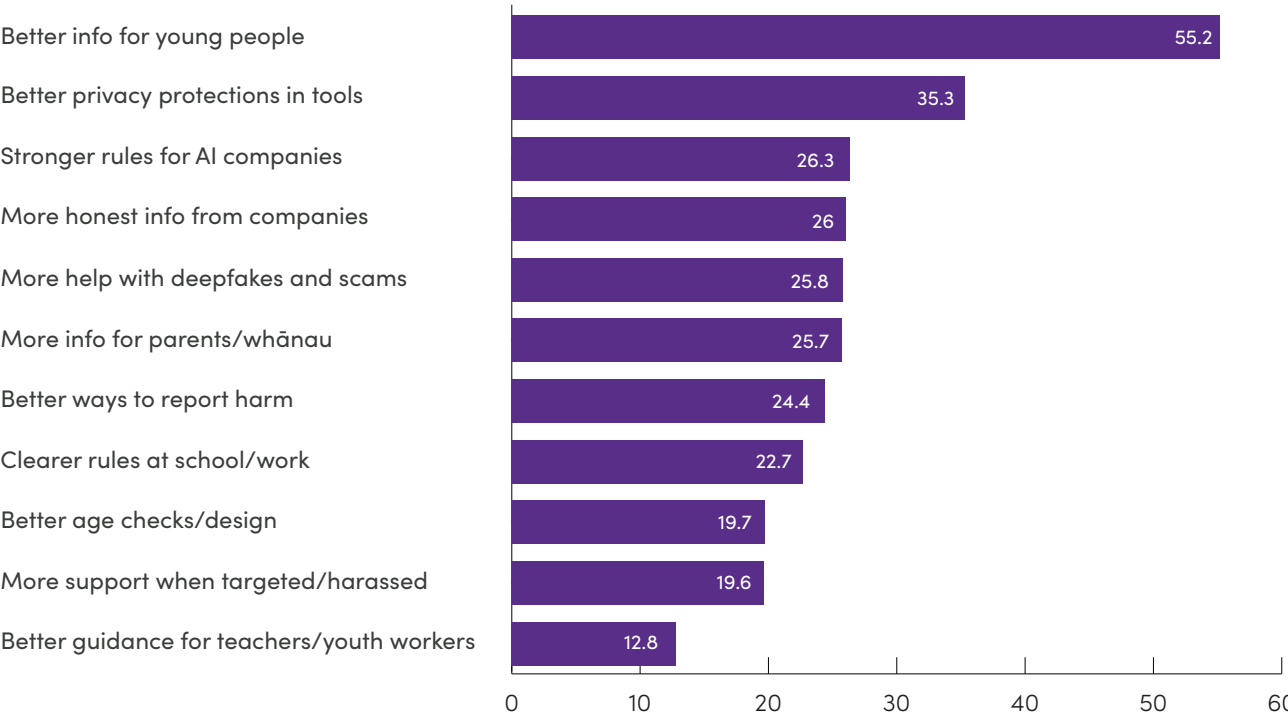
Young people are clear about the type of support they need. The most-wanted support was better, practical information directly for young people (a priority for 55%).

- **Top Information Needs:** Practical privacy checklists, worked examples of “what safe use looks like,” and guidance on navigating deepfakes and scams.
- **Preferred Delivery:** Real stories from other young people, delivered through schools or kura, and short-form social video platforms (TikTok, Instagram, YouTube)

Respondents were also asked to name the biggest priorities right now. They put better information for young people at the top (55%). Other common priorities included better privacy protections in tools (35%), stronger rules for AI companies (26%), more honest information from companies about risks (26%), more help with deepfakes and scams (26%), and more information for parents and whānau (26%).



Figure 13. Respondents’ top priorities for action



The survey’s final open questions invited respondents to tell adults what they most need to understand. Responses centred around a general theme, with respondents identifying AI use is complex, multifaceted, and here to stay – and young people want support, guidance, and trust with navigating it.

“Youth using AI is inevitable and should not be stopped. What should be done is guide them down the uncertain path, not to coddle but to prepare them.”

The free text also highlighted that understanding why young people use AI is crucial to offering support. This is an important insight given the reported use of AI for emotional support, information-seeking in the absence of trusted adult support, and assistance with schoolwork pressure.

“We don’t use it to cheat or to get answers. Some might, majority just use it to get ideas and build off of that. If we could make it so there isn’t a stigma around the use of AI and teach kids how to use [it] [safely]...”

“I think the risks are not anywhere near well known enough.”

4.13 Discussion, limitations and conclusion

The survey findings offer insight into how young people in Aotearoa New Zealand are engaging with generative AI. While young people are early adopters of a wide range of AI tools, they also want more support, guidance, and trusted adults to help them navigate this rapidly evolving technology.

Overall, the findings suggest that young people perceive AI to be a valuable and increasingly integrated part of their digital lives. Most respondents used AI primarily for study and learning, valuing it for finding information, explaining concepts, and supporting academic tasks. Around a third used it for information searching, brainstorming, fact-checking, planning, and everyday questions, highlighting that AI is becoming embedded well beyond education. Respondents also reported AI image, video, and audio generators, AI companions, and AI-generated sexual content. Although only 2.6% reported using AI companions – substantially lower than international findings (Robb & Mann, 2025) – this difference may reflect the younger age profile of the survey sample.

The qualitative responses indicated that many young people engage thoughtfully with AI, recognising its strengths while remaining aware of its limitations. However, the findings also highlight several emerging risks. Sixty per cent of respondents reported encountering incorrect information, an important finding given AI may increasingly become young people's first point of reference for information. Another notable finding was that 30% of young people used AI for emotional support or advice. Respondents commonly described AI as providing a non-judgemental space and cited fear of talking to adults as a reason for turning to AI.

"It felt easier than talking to a person."

"I did not want to be judged."

Reliance on AI also emerged as a recurring concern, with "becoming too reliant on an AI tool" the most commonly reported personal impact. Other reported harms included incorrect information and misleading media, exposure to violent and sexual AI-generated content, and pressure to continue engaging. As AI becomes increasingly embedded across everyday digital platforms, these findings reinforce the importance of helping young people develop safe, balanced, and critical approaches to AI that complement, rather than replace, trusted human relationships and sources of support.

Deepfakes also emerged as a significant issue, with around one in nine respondents reporting they had either been targeted by a deepfake or knew someone who had. Although this survey did not examine help-seeking, international evidence indicates many victims do not disclose these experiences or seek support. Combined with the finding that 12.8% had encountered AI-generated NSFW content, this reinforces the importance of accessible, shame-free education and support around deepfakes and AI-generated sexual content.

Privacy and safeguarding also emerged as important areas of concern. Many respondents reported entering substantial amounts of personal information into AI tools despite having limited understanding of how that information is stored, shared, or used. Young people consistently identified a need for clearer guidance on protecting their privacy, alongside stronger safeguards within AI platforms.

These findings should be interpreted alongside several limitations. The survey used a self-selecting convenience sample rather than a probability sample, so prevalence estimates should be considered directional rather than representative. The sample was skewed towards younger respondents, females, and participants from Wellington and Auckland, with evidence of school-based clustering, limiting subgroup comparisons. Findings are also based on self-report, and a small number of questions intended as multi-select were presented as single-select, reducing response detail. In addition, some respondents expressed strongly critical views about AI, which may have influenced the framing of some open-text responses.

Despite these limitations, the survey provides an important youth-centred evidence base that aligns closely with international research and frontline practitioner experience. Overall, the findings suggest that supporting young people to use AI safely does not require discouraging its use. Rather, it requires equipping them with the knowledge, critical thinking skills, privacy awareness, and trusted adult support needed to engage with AI confidently, safely, and responsibly. This convergence with international evidence is explored further in the following section.

5. Overall Summary



Generative AI is rapidly becoming part of young people's everyday lives, shaping how they learn, connect, and seek support. Drawing together international and local research, frontline service feedback, and the views of young people in Aotearoa New Zealand, this project examined the evolving landscape of generative AI use among young people, exploring how and why they use it, the risks and harms they encounter, and what support they want to navigate it safely.

Overall, several key themes emerged:

- **AI is now a routine part of many young people's everyday lives.** Generative AI is used widely and frequently by young people for a broad range of purposes.
- **Young people see both value and risk in AI.** While valuing AI, many young people remain concerned about its impacts on jobs, the environment, and creative or critical thinking.
- **Risks and harms are widespread, while digital safeguarding is limited.** Young people encounter a range of risks while using AI, with the risk of reliance emerging as a particular concern. These risks are compounded by inconsistent privacy and safeguarding habits.
- **AI is increasingly used for emotional support.** Both locally and internationally, young people are turning to generative AI for advice, guidance, and emotional support.

- **AI companions are an emerging concern.** AI companion use is growing rapidly among young people, but can be associated with significant risks, including emotional dependency, harmful advice and information, and exposure to sexualised content.
- **AI-generated sexual imagery and deepfakes present growing risks to young people.** Exposure to deepfakes is increasing alongside the rapid expansion of AI-generated sexual content and platforms, most of which operate with minimal safeguards and regulation.
- **AI is outpacing guidance and support.** Adults and professionals consistently report that the rapid evolution of AI is outpacing their knowledge and ability to provide support. Alongside this, many young people won't turn to adults if something goes wrong, due to fear of getting into trouble, embarrassment, and a belief that adults don't understand AI.

As AI continues to evolve, strengthening young people's capability to engage with it safely, alongside frontline services' capacity to support them through this, is critical. Key priorities include equipping young people with AI literacy and safeguarding skills; strengthening frontline services' understanding of generative AI, including how to respond to experiences of harm; and pursuing broader platform and policy changes. The recommendations that follow set out how this can be achieved.

6. Recommendations



These recommendations are drawn from all three evidence sources and are framed to drive the next phase. They are grouped by audience. Each is anchored in findings reported above.

6.1 For the resources themselves (Phase Two)

1. **Lead with real stories from young people.** This was the clear format preference. Build resources around authentic rangatahi voices and scenarios rather than top-down rules. Where youth advisory groups already exist, such as Netsafe's Youth Action Squad, these should be used to co-design AI safety resources and test emerging issues.
2. **Prioritise privacy checklists and "what safe use looks like."** These were the two most-requested content types, and they map onto the inconsistent privacy behaviours observed.
3. **Make text-detection, fraud detection and verification a core skill.** Confidence was lowest for AI text, the most-used modality. Teach concrete habits for checking AI text and claims, alongside spotting fake images. Further, given the growing prevalence of AI-enabled scams and fraud, preparing young people for the age of AI-enabled fraud and deception is a core priority.
4. **Address reliance and emotionally persuasive AI, including companions, explicitly and without shame.** Reliance was the top personal harm, and the frontline and research both flag companions, alongside a wider and evolving category of emotionally adaptive chatbots and synthetic relationships. Include guidance on healthy boundaries, building critical thinking, and test companion messaging locally given the survey divergence.
5. **Cover deepfakes and synthetic intimate imagery directly.** With around one in nine reporting proximate or direct contact, and frontline reports of nudify-app harm, name the issue, explain the law, offer resources to build critical thinking and ethical online engagement, and give low-stakes reporting pathways.
6. **Design for low bandwidth and for inclusion.** Many respondents face access barriers. Offer lightweight formats, a te ao Māori perspective, and inclusive framing for rainbow, takatāpui and neurodivergent young people.

6.2 For parents, caregivers and whānau

1. **Equip the adults young people actually approach.** Whānau are the leading first port of call, so plain-language guides for them will have outsized reach.
2. **Coach a calm, curious response.** Fear of overreaction and of getting in trouble suppress help-seeking. Offer resources on how to have safe, shame free and mana-enhancing conversations with young people, so young people feel safe to seek support and disclose harm.
3. **Offer tools for building youth resiliency and protective factors.** Support whanau with meaningful and digitally empowering approaches to equipping young people around AI, including fostering protective factors and building resilience.

6.3 For schools, kura and the frontline workforce

1. **Make rules clear and de-stigmatising.** A third of respondents lack clear, understood rules, and ambiguity fuels secrecy. Frame reasonable-use policies around learning, not policing.
2. **Invest in workforce capability.** Frontline services report feeling out of their depth (The Light Project, 2026). Fund training, prevention and assessment tools, and response frameworks specific to AI-related harms, including companions and synthetic intimate imagery.
3. **Lower the barrier to formal help, and meet young people on their channels.** Signpost confidential pathways, and pair school delivery with short-form social video and trusted creators.

6.4 For platforms

1. **Strengthen privacy-by-default and transparency.** Respondents called for honest information from companies and better default protections; the testing literature shows these are not the norm (AISF, 2025).
2. **Safeguard against synthetic media of real people,** with accessible reporting and takedown routes, reflecting both the survey and the research on the gendered impact of AI-generated deepfakes (McGlynn et al., 2026).
3. **Adopt Safety by Design, including meaningful age assurance and content moderation.** Testing found a lack of effective age-gating across popular agents and companion products, including those marketed to children (AISF, 2025; eSafety Commissioner, 2025).

6.5 For government and policymakers

1. **Develop a coordinated national approach to AI safety education,** aligning schools, families, industry and frontline services, and recognising that responsibility for safe AI use cannot rest on young people alone.
2. **Pursue age-appropriate, Safety by Design standards,** drawing on emerging codes overseas (eSafety Commissioner, 2025) and adapting them to the Aotearoa context.
3. **Clarify and communicate the legal position on synthetic intimate imagery involving under-18s,** which constitutes child sexual exploitation, given the sharp rise in school-based cases reported to Netsafe (Netsafe, 2026).
4. **Resource the frontline and education workforce.** Services report feeling out of their depth (The Light Project, 2026); the training and response frameworks recommended in 6.3 need a funder.
5. **Invest in the local evidence base.** Fund a follow-up, more representative survey with quotas across age, gender, ethnicity and region, to track change and resolve the companion divergence identified in Section 5. Longitudinal research, following the same young people over time, would add particular value in tracking how attitudes, risks and use patterns evolve over time. Invest in longitudinal approaches involving youth and AI safety.
6. **Build early warning capability.** New Zealand currently lacks a dedicated capability to monitor AI-related harms as they emerge. Building a national AI harms early warning capability combining frontline reporting, youth surveys and international horizon scanning, would help identify emerging harms before they become widespread (Netsafe, 2026).

6.6 For researchers

1. **Build a more representative picture of young people's AI use,** with quotas across age, gender, ethnicity and region, and particular attention to Māori, Pacific and disabled young people, who are underrepresented in this survey and in the wider evidence base.
2. **Resolve the open questions this survey raises,** including the AI companion divergence identified in Section 5.2 and what young people mean by becoming "too reliant."
3. **Evaluate the Phase Two resources** once deployed, so the next phase is tested rather than assumed.

7. References

- Artificial Intelligence Safety Forum (AISF). (2025). **Generative AI Safety Reports: General-Purpose Conversational AI Agents** (April 2025; October 2025); **AI Companions for Teens and Adults** (April 2025); **AI Companions for Children** (September 2025). safetyforum.ai
- British Association of Counsellors and Psychotherapists. (2025). **“Therapists warn of dangers as children turn to AI for mental health advice”**.
- Clark, A. (2025). **The Ability of AI Therapy Bots to Set Limits With Distressed Adolescents: Simulation-Based Comparison Study**. JMIR Mental Health, 12, e78414.
- Common Sense Media. (2026). **The Common Sense Media Census: AI Use by Tweens and Teens**. Common Sense Media, Youth AI Safety Institute.
- eSafety Commissioner. (2025). **AI companions: information sheet for educators**. Australian Government eSafety Commissioner.
- Internet Watch Foundation (2026). **“AI becoming ‘child sexual abuse machine’ adding to ‘dangerous’ record levels of online abuse, IWF warns”**.
- Lin, Z. (2025). **Functional intersubjectivity: Toward a well-being framework for AI companions**. 10.31234/osf.io/rxedj_v1.
- Lyons, A., Young, J., Sing, F., Goodwin, I., McCreanor, T., Moewaka Barnes, A., & Huckle, T. (2025). **“Almost constant”: Young people overwhelmed by harmful online marketing**. Public Health Communication Centre Aotearoa.
- McGlynn, C., McDermott, Y., Macdonald, S., Toparlak, R. T., Tarrant, F., & Treacy, S. (2026). **Invisible No More: How AI Chatbots Are Reshaping Violence Against Women and Girls**. Durham University, Swansea University, UK Research and Innovation.
- Netsafe. (2026). **Generative AI in online harm reported to Netsafe: a focus on young people**. Netsafe, Aotearoa New Zealand.
- Netsafe (2025) **AI and online safety**. Available at: https://resource.netsafe.org.nz/Final-version--Netsafe_AI-API-wCVR_121224R3.pdf (Accessed: 17 July 2026).
- Robb, M. B., & Mann, S. (2025). **Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions**. Common Sense Media.
- The Light Project. (2026). **Feedback from frontline workers** (Jan 2025 – June 2026).
- Thorn. (2025). **Deepfake Nudes & Young People: Navigating a new frontier in technology-facilitated nonconsensual sexual abuse and exploitation**.
- Zhang, R., Li, H., Meng, H., Zhan, J., Gan, H., & Lee, Y. C. (2025, April). **The dark side of ai companionship: A taxonomy of harmful algorithmic behaviors in human-ai relationships**. In Proceedings of the 2025 CHI conference on human factors in computing systems (pp. 1-17).

Note: the sources informed the introduction and overall summary.

Appendix

Sources at a Glance

All bases for the youth survey are N = 1,202 unless noted.
The table summarises the supporting evidence used in the introduction and summary.

Source	Origin	Focus
Youth survey (this report)	Aotearoa New Zealand	1,202 young people 13–24: use, risks, support needs
The Light Project (2026)	Aotearoa New Zealand	Frontline practitioner feedback on AI-related harms
Common Sense Media (2026)	United States	1,204 children 9–17: AI use, dependency, safety gaps
Robb & Mann (2025)	United States	1,060 teens 13–17: how and why teens use AI companions
Clark (2025)	United States	Safety test: AI bots and distressed adolescents
AISF (2025)	Aotearoa New Zealand	Independent safety ratings of agents and companions
eSafety Commissioner (2025)	Australia	AI companions: risks and educator guidance
McGlynn et al. (2026)	United Kingdom	Chatbots and violence against women and girls
Lyons et al. (2025)	Aotearoa New Zealand	Young people and harmful online marketing

About the Authoring Organisations



The Artificial Intelligence Safety Forum (AISF) is a New Zealand-based nonprofit, self-regulatory forum for safety in products that use generative AI. safetyforum.ai



The Light Project is a workforce development service that supports youth professionals and communities to equip young people to positively navigate their digital sexual environments and minimise the risk of online harm. thelightproject.co.nz



Netsafe is an independent non-profit organisation supporting people in Aotearoa New Zealand to have safe and positive online experiences. We keep people of all ages safe online by providing free support, advice, and education. netsafe.org.nz