

Consultation response form

Your response

Question	Your response
The incidence of content that is harmful to children on regulated services We invite stakeholders to provide evidence relating to the incidence of PPC, PC or NDC on regulated user-to-user, search or combined services. Please provide as much detail as you can. This could include but is not limited to evidence about:	
The quantity, prevalence or presence of specific kind(s) of PPC, PC or NDC on regulated services and the frequency with which children are encountering this content.	Confidential? – N Children engaging with chatbots and content algorithms directed by artificial intelligence are encountering harmful online content. Studies on susceptibility to misinformation demonstrate that we all engage with new information differently according to psychological dispositions like emotional well-being, loneliness, and open-mindedness.¹ Even social media platforms designed to interact with other humans expose children to automated recommendations that can introduce them to personalised AI-generated material. Given the exponential boom in large language models, children are becoming increasingly bombarded with emotionally-

¹ Dustin P. Calvillo et al., 'Personality Factors and Self-Reported Political News Consumption Predict Susceptibility to Political Fake News', *Personality and Individual Differences* 174 (May 2021): 110666, <https://doi.org/10.1016/j.paid.2021.110666>.

Question	Your response
	manipulative and unreliable messaging that lacks the guardrails of conventional forms of media. Current psychological research into misinformation indicates that social isolation (an exacerbated factor post-pandemic) and a lack of self confidence predict susceptibility to AI gateways into dangerous content that can trigger unhealthy habits, bullying, and self-harm.
How demographic characteristics such as age, gender, race, sexual orientation and disability influence the likelihood that children will encounter specific kind(s) of PPC, PC or NDC.	Confidential? – N Psychological testing has shown that people can process misleading or harmful information differently depending on their social demographics. The youth have been shown to be most at risk due to a general increase in anxiety, depression, and social pressure as well as an underdeveloped cognitive ability to discern credibility. Teenagers already living with emotional distress, poor mental health, and social isolation may have a higher chance of interfacing with AI companions as a form of socialisation and affirmation not provided at school or in the home. On top of age, variables like gender and personality profiles can also shape engagement. For example, individuals who are more agreeable have been shown to trust conversational chatbots that simulate real human interaction. Children exhibiting neurodivergent traits may be particularly susceptible to unhealthy communication with AI, especially those lacking proper outside resources and responsible parties to monitor activity. These circumstances tend to increase one's likelihood of encountering dangerous content rendered by artificial intelligence.
The services or types of service where children are most likely to encounter specific kind(s) of PPC, PC or NDC.	Confidential? – N Children are most likely to find dangerous and artificial messaging via algorithmic recommendations, interactive

Question	Your response
	chatbots, and customized digital environments on social media. The services these platforms provide prioritise continued engagement above other goals such as mental health or accurate knowledge. AI chatbots (such as Chat-GPT, Claude, Gemini, etc.) may be most harmful given their singular ability to simulate genuine interaction with the user. While traditional media occurs in one direction, these conversational systems adjust to each response in order to give bespoke answers that align with one's ideology and worldview. This ability to satisfy the user's sensibilities renders the technology more convincing as well as more addictive, leading to documented side effects resembling drug use. These relatively new features create conditions where already at-risk members of the population depend on AI as an authority, despite its proven lack of accuracy.
The severity of harm that UK children suffer, or may suffer, as a result of content that is harmful to children We invite stakeholders to provide any evidence regarding the severity of harm that UK children suffer, or may suffer, as a result of encountering PPC, PC or NDC. Please provide as much detail as you can. This could include but is not limited to evidence about:	
The impact that specific kind(s) of PPC, PC or NDC can have on a child's wellbeing, behaviours, or other physical or psychological outcomes. This might include evidence concerning cumulative exposure to specific kind(s) or combinations of PPC, PC or NDC.	Confidential? – N Studies have demonstrated that messaging becomes more believable when exploiting the emotions of the recipient. Adolescents who assign emotional-importance to their relationship with a simulated conversational system may mistake it for an actual form of socialisation or source of expertise. Psychological research into parasocial

Question	Your response
	relationships suggests that mediated representation or digital presence can create one-sided but seemingly authentic relationships, extending even to parties that are not alive or real. When projected onto chatbots, parasocial relationships can cause young people to re-establish their hierarchy of needs, replacing actual socialisation with a device trained to emotionally validate the user in a manner that feels trustworthy. Overdependency on this technology has been shown to decrease one's ability to emotionally self-regulate thereby perpetuating usage at the expense of autonomy. This development becomes all the more troubling when it occurs at a young age when children first experiment with social independence.
Any causal links between a child encountering specific kind(s) of PPC, PC or NDC and subsequently suffering physical or psychological harm.	Confidential? – N Current academic studies into adolescent chatbot engagement have demonstrated that the AI exploits the users emotions in dangerous ways. Those most susceptible to maladaptive use of these computer programs may encounter responses that reinforce harmful thought patterns, affirm delusional spirals, or simply provide misinformation on issues of significance. Recent publicised incidents involving unhealthy relationships between AI chatbots and young users illustrate the real dangers of this new technology. Although more studies are necessary, the potential for irrational persuasion and manufactured intimacy indicates opportunities for AI-generated content to cause personal harm.

Question	Your response
How demographic characteristics such as age, gender, race, sexual orientation and disability can influence the severity of harm that UK children suffer, or may suffer, as a result of encountering specific kind(s) of PPC, PC or NDC.	Confidential? – N Several studies have identified age’s negative correlation with conspiracy theory susceptibility, a trend Fox and Williams attribute to the decline in stress one feels while aging.² An apparent naiveté renders younger demographics more gullible to misleading scams, particularly in the online domain.³ This only compounds in the age of generative AI. Some publications have identified women as particularly at risk for believing in and sharing misinformation while other studies have designated men as the most likely party to endorsing conspiracy theories.⁴ While there is less research in regard to trust in LLMs, one Norwegian survey of 2692 higher education students found that women demonstrated more hesitation to trusting chatbots.⁵ Multiple surveys targeting the patronage of Replika and other companion bots found the majority to be male.⁶
Whether it is appropriate to make changes to the kinds of PPC and PC in the Act	

² Fox and Williams, ‘Do Stress and Anxiety Lead to Belief in Conspiracy Theories?’

³ Sarno et al., ‘Which Phish Is Captured in the Net?’

⁴ Jinhui Li and Xiaodong Yang, ‘Does Exposure Necessarily Lead to Misbelief? A Meta-Analysis of Susceptibility to Health Misinformation’, *Public Understanding of Science* (Bristol, England) 34, no. 2 (2025): 222–42, <https://doi.org/10.1177/09636625241266150>; Erin C. Cassese et al., ‘Gender Differences in COVID-19 Conspiracy Theory Beliefs’, *Politics & Gender* 16, no. 4 (2020): 1009–18, <https://doi.org/10.1017/S1743923X20000409>.

⁵ Anja Møgelvang et al., ‘Gender Differences in the Use of Generative Artificial Intelligence Chatbots in Higher Education: Characteristics and Consequences’, *Education Sciences* 14, no. 12 (2024), <https://doi.org/10.3390/educsci14121363>.

⁶ Vivian Ta et al., ‘User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis’, *Journal of Medical Internet Research* 22, no. 3 (2020): e16235, <https://doi.org/10.2196/16235>; Auren R. Liu et al., ‘Chatbot Companionship: A Mixed-Methods Study of Companion Chatbot Usage Patterns and Their Relationship to Loneliness in Active Users’, arXiv:2410.21596, version 3, preprint, arXiv, 11 August 2025, <https://doi.org/10.48550/arXiv.2410.21596> (though see Michelle Faverio and Olivia Sidoti, ‘Teens, Social Media and AI Chatbots 2025’, *Pew Research Center*, 9 December 2025, <https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/>).

Question	Your response
We invite stakeholders to provide evidence that suggests any kind(s) of PPC and/or PC in the Act are inappropriate or incorrectly categorised. This could include but is not limited to evidence indicating that:	
Any kind of PPC or PC in the Act should cease to be categorised as content harmful to children.	Confidential? – N The literature reviewed makes no meaningful suggestion that any harmful content categories of should be removed. Instead, it indicates that novel forms of harm resulting from generative AI systems remaining outside the scope of existing categories.
Any kind of PPC would be more appropriately categorised as PC.	Confidential? – N Current studies do not provide sufficient evidence to justify the reclassification of existing categories from PPC to PC. However, the increasing concern over AI-generated content indicates that some material previously unclassified as harmful may now be exhibiting a larger influence.
Any kind of PC would be more appropriately categorised as PPC.	Confidential? – N The current categorical system functions appropriately as it stands, however, new media including interactive large language models have now disproportionately put children at risk due to the emotionally manipulative nature of the responses. Unregulated, this technology can prove significantly more harmful to children than traditional online content.
We invite stakeholders to provide evidence in support of additional kind(s) of content being added to the	

Question	Your response
PPC or PC categories in the Act. This should include, but is not limited to:	
Evidence that the kind(s) of content pose a material risk to an appreciable number of children.	Confidential? – N New research on AI-generated content observes large language models “hallucinating” and producing inaccurate information under the facade of credibility. One study found that ChatGPT endorsed incorrect statements up to 26% of the time depending on the prompt.⁷ As more and more young adults adopt AI-mediated platforms, exposure to misinformation will likely spread As the use of AI chatbots increases among younger users, exposure to harmful content will spread, in large part due to the accessible, friendly, and professional branding curated by companies like Anthropic, Open.AI, and Google. Researchers have documented harmful side effects from these products with one study correlating trust in chatbots with psychological dependence, yielding salience, tolerance and withdrawal, a marker for addiction.⁸
Evidence that the kind(s) of content pose a risk of significant harm to children.	Confidential? – Y / N Large language models using artificial intelligence pose greater threats than misinforming the youth. By mimicking compassion and unconditional support, function as confidants without the appropriate guardrails that real adult humans use when engaging with children, particularly those in distress. We have already seen mentally ill young chatbot users develop problematic relationships with AI only for it to

⁷ Aisha Khatun and Daniel G. Brown, ‘Reliability Check: An Analysis of GPT-3’s Response to Sensitive Topics and Prompt Wording’, arXiv:2306.06199, preprint, arXiv, 9 June 2023, <https://doi.org/10.48550/arXiv.2306.06199>.

⁸ Tianling Xie et al., ‘Friend, Mentor, Lover: Does Chatbot Engagement Lead to Psychological Dependence?’, *Journal of Service Management* 34 (May 2023): 806–28, <https://doi.org/10.1108/JOSM-02-2022-0072>; Mohammed Salah et al., ‘Chatting with ChatGPT: Decoding the Mind of Chatbot Users and Unveiling the Intricate Connections between User Perception, Trust and Stereotype Perception on Self-Esteem and Psychological Well-Being’, *Current Psychology* 43, no. 9 (2024): 7843–58, <https://doi.org/10.1007/s12144-023-04989-0>.

Question	Your response
	encourage suicide and violence towards other, in some cases successfully so.⁹ Persistent trust in AI chatbots may lead children to harmful forms of misinformation or echo pre-existing negative cycles of thought. This becomes increasingly detrimental when AI is used for personal information or reflection, which can cause distorted self-perception that can lead to irrational decision-making.
Evidence to support each kind of content being categorised as either PPC or PC.	Confidential? – N The use of AI to generate misinformation and unhealthy chatbot relationships may be most appropriately classified as Priority Content (PC) because they pose a legitimate threat of psychological harm or behavioural influence. That said, specific types of AI interaction may indeed qualify as Priority Primary Content (PPC), particularly when they provide customised responses taken as real-world knowledge, especially in fields like health. It is generative AI's ability to interact in such a personal manner that ultimately persuades its users to believe misinformation and for parasocial attachments.

Please tell us how you came across about this call for evidence.

- ☐ Email from Ofcom
☒ Saw it on social media

⁹ Blake Montgomery, 'Mother Says AI Chatbot Led Her Son to Kill Himself in Lawsuit against Its Maker', Technology, *The Guardian*, 23 October 2024, <https://www.theguardian.com/technology/2024/oct/23/character-ai-chatbot-sewell-setzer-death>; Bobby Allyn, 'Lawsuit: A Chatbot Hinted a Kid Should Kill His Parents over Screen Time Limits', Business, *NPR*, 10 December 2024, <https://www.npr.org/2024/12/10/nx-s1-5222574/kids-character-ai-lawsuit..>

- ☐ Found it on Ofcom's website
- ☐ Found it on another website
- ☐ Heard about it on TV or radio
- ☐ Read about it in a newspaper or magazine
- ☐ Heard about it at an event
- ☐ Somebody told me or shared it with me
- ☐ Other (please specify)

Please complete this form in full and return to ContentHarmfultoChildren@ofcom.org.uk.