

Is Instagram now safe for teens?

Instagram Teen Accounts Case Study

About 5Rights Foundation

5Rights develops new policy, creates innovative frameworks, develops technical standards, publishes research, challenges received narratives and ensures that children's rights and needs are recognised and prioritised in the digital world. While 5Rights works exclusively on behalf of and with children and young people under 18, our solutions and strategies are relevant to many other communities.

Our focus is on implementable change and our work is cited and used widely around the world. We work with governments, intergovernmental institutions, professional associations, academics, businesses, and children, so that digital products and services can impact positively on the lived experiences of young people.

APRIL 2025

Introduction

WARNING – This case study contains screenshots of sexualised imagery, and toxic comments that some could find upsetting.

This case study reflects preliminary findings from ongoing deeper research into Instagram’s Teen Accounts – introduced by Meta in late 2024 –, among other services. While the changes were a step towards teen safety on the platform, our testing aimed to assess the everyday experience of teenagers, using avatar accounts that mirrored casual use.

The early stages of this investigation revealed that young users signing up to new accounts are still exposed to risky contact recommendations, hidden advertising, sexualised or misleading content and addictive design patterns. The investigation also identified limitations in transparency and accessibility of governance documents, as well as ineffective and inconsistent age assurance.

Background

At 5Rights, we [welcomed the shift towards higher privacy settings](#) for minors, while warning that more needed to be done to ensure teens feel truly safe on the app. In particular, the increased focus on parental surveillance rather than tackling existing harms in Instagram’s pervasive profiling and recommendation system risked limiting children’s privacy rather than enhancing it. Additionally, it remained unclear how effectively these new privacy settings limit the unnecessary collection, sharing and commercial exploitation of children’s data by Meta.

In September 2024, Meta announced the introduction of Instagram Teen Accounts, [describing them as a major advancement in protecting young users online](#). These accounts were promoted as having different default privacy settings, notification ‘quiet times’ between 10pm and 7am, ‘break’ reminders, and the option for parental controls. These features are a step towards better safety. The issue is they are not enough.

This case study provides some insights into whether Instagram’s protections deliver meaningful change in practice, analysing the experience of teenage users when they sign up for a new Instagram account. Given that nearly half of 8-17-year-olds in the UK and 62% of US teenagers use Instagram ([Ofcom](#), [Pew Research Center](#)), the stakes for children’s safety on the platform are high. Despite Meta’s claims of prioritising teen safety, our findings show that minors remain exposed to inappropriate content, conduct, contact and commercial risks.

Method

From January to April 2025, three researchers conducted semi-structured, real-world testing of Instagram's Teen Accounts, as part of the preliminary phase of a deeper investigation. This included analysing public documents such as Instagram's Terms of Service and creating three 'avatar' accounts on separate devices. These accounts were designed to mimic the natural journey of a young person signing up for Instagram, without changing default privacy settings or deliberately seeking out edge-case content.

To simulate 'realistic' teen experiences, avatars had generic usernames and neutral, age-appropriate profile photos (e.g. a cartoon or default icon). Each was assigned a date of birth placing them between 14 and 17, along with a common name. During the default onboarding, we selected broad, mainstream interests like music, fashion, and sport – avoiding anything extreme or niche. Our goal was to mirror casual, everyday use rather than test out high-risk cases.

We observed how the platform responded to standard user interactions: liking a random selection of suggested posts, following a random selection of recommended accounts, and scrolling through the home feed and recommendations. We also experimented with using features like Direct Messaging, Reels and Stories. Rather than replicating highly specific or rigid activity patterns, we allowed for variation in behaviour to reflect a wider range of possible user pathways.

The methodology used is designed to be evidence-based, adaptable and outcome-oriented, using a structured set of questions. Further research is ongoing.

Limitations

We acknowledge the following limitations to this preliminary study:

- This testing was constrained by the number of available devices, time, and research team capacity. While these findings reflect what *can* reasonably happen while using Instagram's Teen Accounts, they do not claim to reflect what *always* happens.
- Our methodology is intended first and foremost to support compliance investigations, placing more emphasis on testing against clear obligations, over the full range of borderline experiences that may fall outside of existing regulations.
- The devices used in this case study are also part of broader avatar investigations across other services. While efforts were made to prevent cross-platform influence (e.g. clearing cookies, disabling app tracking), we cannot fully rule out the possibility of external factors affecting algorithmic recommendations.

In presenting our preliminary findings, some names, usernames and faces of non-public figures have been censored to preserve their privacy.

Our preliminary findings

Our testing reveals that Instagram’s Teen Accounts present concerns across several key areas.

Ineffective age assurance

One of the most significant issues is Instagram’s ineffective age assurance system. Meta (Instagram’s parent company) claims that users must be at least 13 years old to sign up for an account. The company states that they [verify the user’s age using a variety of methods](#), including providing a date of birth, photo identification, or submitting a video selfie. In practice however, age assurance is weak or inconsistently implemented. All our avatars were able to register using only self-reported birthdates, with no additional checks.

Risky contact recommendations

Upon signing up, Instagram immediately suggested to our avatar accounts to connect with adult-owned accounts. While some restrictions exist – such as preventing adults from messaging minors unless they follow each other – our testing found that teen avatars could also still freely send message requests to adults. In addition, the algorithm routinely encouraged our avatar accounts to widen their network of contacts with unknown strangers, increasing potential privacy and safety risks.



Figure 1 - Algorithmic connection recommendations, including to adult-owned accounts, upon sign-up.

Nudges for more data

Our testing found that Instagram nudges teens to add their full names to their profiles, with the justification that it helps ‘friends know it’s you’. In some cases, our avatar accounts received follow requests immediately upon signing up, suggesting that our teen avatars’ profiles are discoverable by default. This raises concerns about the potential for these accounts to be targeted by unknown individuals, including bad actors. In addition, avatars were also encouraged to pair their Facebook and Instagram accounts, leading to potential cross-platform profiling.

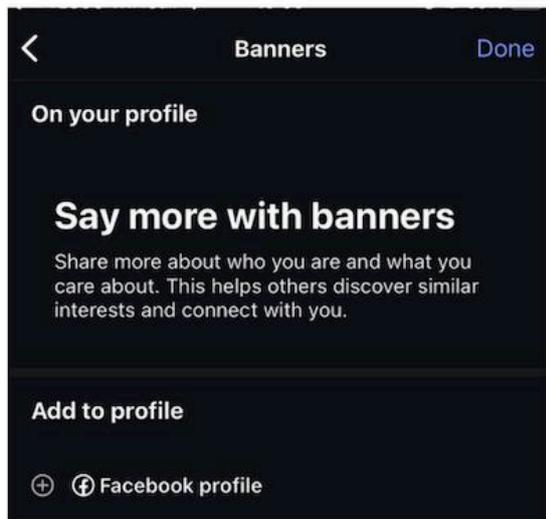


Figure 2 – A prompt to link to a Facebook account.

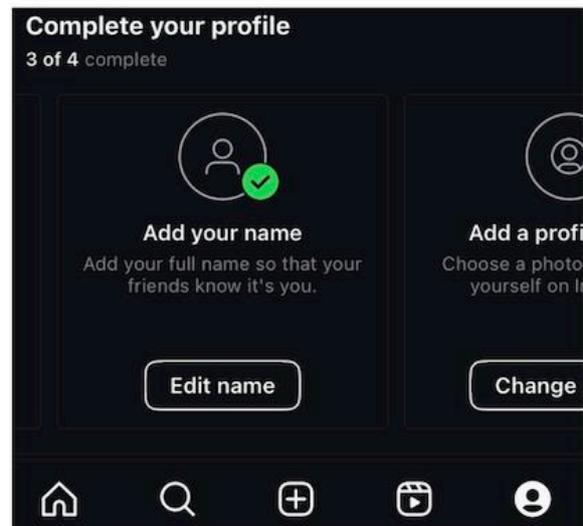


Figure 3 – A nudge to add a full name to our Teen Account.

Persuasive design

During our testing we did not note any meaningful changes to design elements aimed at increasing engagement and time spent. Although there are now screen time reminders, our avatars remained subject to persistent notifications, autoplay, infinite scrolling, and prompts to interact with suggested content. Even after logging off, our accounts received app notifications and emails prompting us to re-engage with the app.

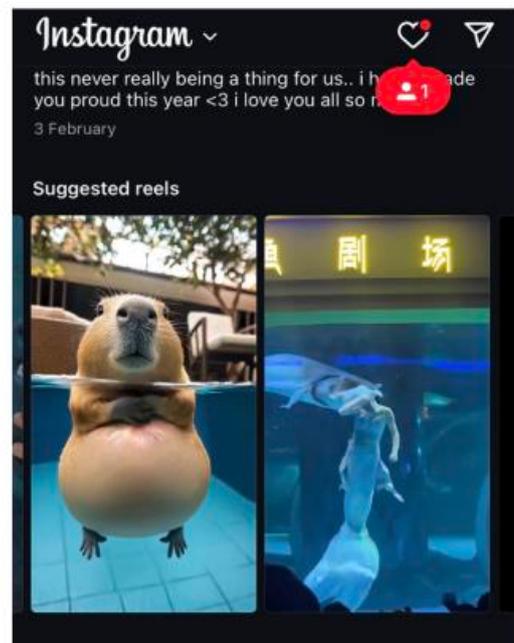


Figure 4 - Suggested reels are integrated into the main feed and open an endlessly scrolling sub-feed of short form video content. This Reels feed is algorithmically refined based on user interactions, which could potentially lead to 'echo chamber' or 'filter bubble' effects.

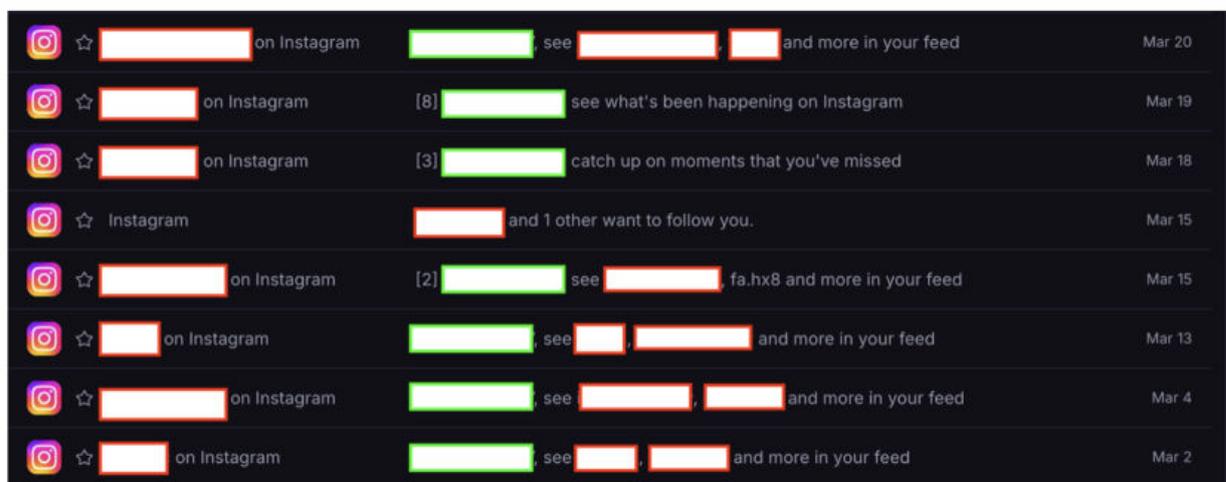


Figure 5 - Constant email nudges to 'catch up'. In many cases, these messages were presented as coming from specific user accounts, though the emails were automated and from Meta. Green boxes in the figure highlight where our own account name was inserted, illustrating how the emails were personalised.

Unclear and inappropriate advertising

From the moment our avatar accounts were created, we observed consistent exposure to influencer advertising and product placement on Instagram. These were often presented in ways to resemble organic posts, making it difficult to distinguish advertising from genuine content. In some cases, ad disclaimers were either buried deep at the bottom of a caption or absent altogether, despite the clear promotional message. Recommended product placements included age-inappropriate products such as lingerie presented with sexualised imagery. These patterns suggest that young people using Instagram may encounter hidden advertising and commercial messaging, without even realising it.

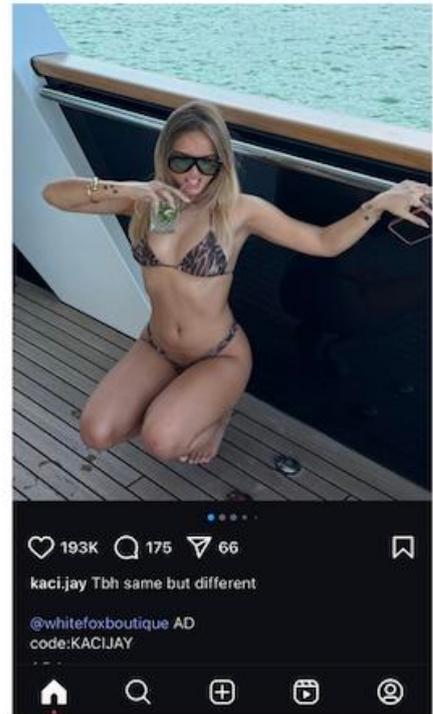


Figure 6 - This celebrity's ad disclosure (served on a 15-year-old male avatar's first session) could not be seen unless the user expanded the full post description.

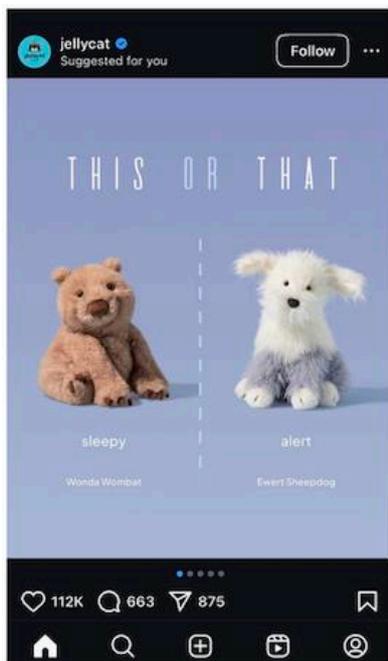


Figure 7 - Corporate messaging without clear advertisement disclosure placed among organic posts through the 'Suggested for You' feature.

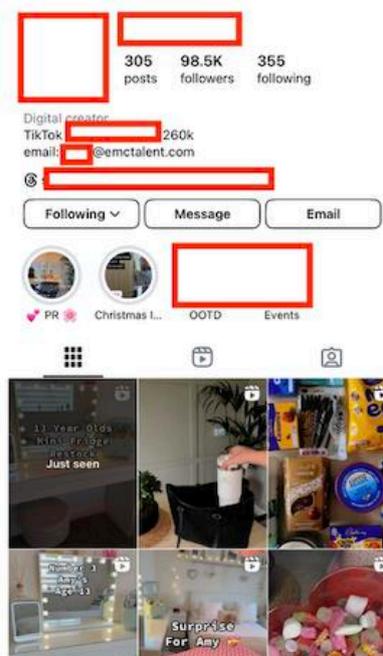


Figure 8 / 9 - This is an example of an account that appears to be devoted to product placements (given the nature of posts and the influencer agency contact details). However, the promotional nature of content is unclear to users - see the example on the left where the caption reads as an advertisement but there is no accompanying disclosure. Many of the account's posts are designed specifically to appeal to teens, using captions suggesting 'hauls' for the account owner's teenage children.



Exposure to sexualised content

In our testing, Instagram did not appear to adequately filter out sexual content, resulting in frequent exposure to sexually suggestive accounts and content for our avatar accounts – despite Teen Account settings. Additionally, sexualised material frequently appeared among the algorithm’s suggested posts and reels. In one instance, a male avatar was recommended a ‘dirty reels’ account focused purely on sharing memes containing sexual innuendos. In another instance, when searching for another avatar’s account to send a follow request, a sex work account with a dubious external link in the profile description was recommended instead.

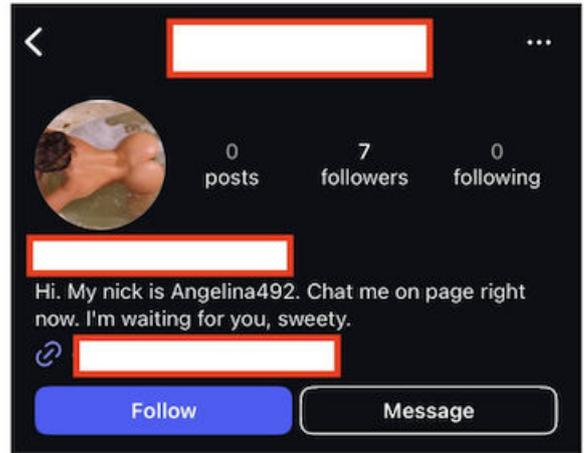


Figure 10 - Dubious account from search results. We reported this account for sexual solicitation, but Instagram responded a week later to say 'it does not go against' Community Standards.

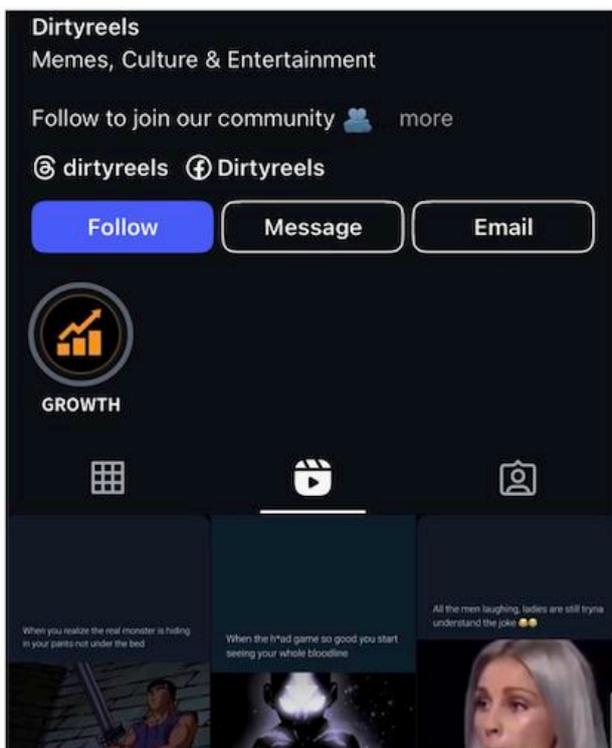


Figure 11 - An account recommended to a male teen avatar through Instagram’s ‘Reels’. Its posts were largely devoted to discussing sex and ‘dirty’ jokes.



Figure 12 - Suggestive content on the main feed.

Amplified misinformation and harmful challenges

During testing, Instagram's 'Discover' function recommended harmful and misleading content to our teen avatars. This included fad weight loss 'challenges' and posts containing medical misinformation.



Figure 13 - A recommended weight loss 'challenge' targeted at one of our female avatars.

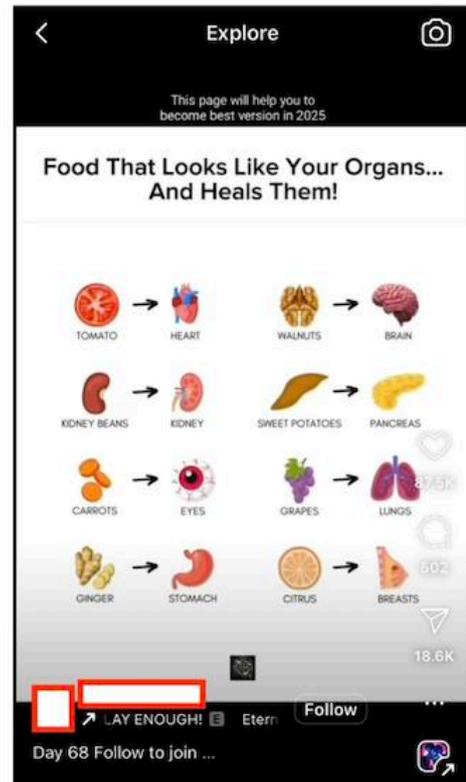


Figure 14 - Unfounded medical misinformation recommended by Instagram's algorithm to a male avatar.

Opaque recommendation systems

Although users are invited to select personal interests when signing up, much of how the recommender system operates remains opaque. For example, avatars were recommended posts focusing on subjects we had never indicated an interest in.

It is unclear what content has been prioritised and why. This lack of clarity leaves young users unaware of how content is being pushed to them and may present skewed views of what is acceptable or popular.

Moderation failures

Our testing revealed frequent failures in content moderation, particularly in the Reels section. Toxic comments were commonly found, targeting individuals on the basis of their ethnicity, body size, gender, and sexual orientation. Despite reporting these hateful comments through the platform, to date no action has been taken by Meta. The only definitive responses we received judged that the content we reported did 'not violate Community Guidelines' and would remain visible. Some examples of the comments can be seen below.



Figure 15 - Toxic comments targeting users based on identity, that according to the timestamp had been visible for a week. It is notable that Meta recently updated their policies to allow users to call some minority groups 'mentally ill' without consequence.

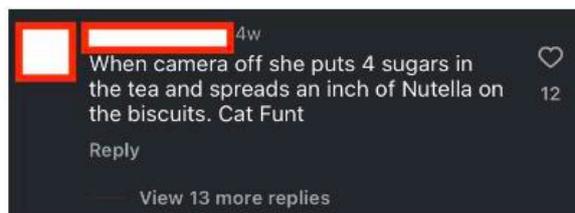


Figure 16 - One post had many comments attacking a woman for her body size. This comment had been visible for 4 weeks and had a variety of other engagements. Other comments suggested the user depicted in the original post may have since quit Instagram due to the harassment received.

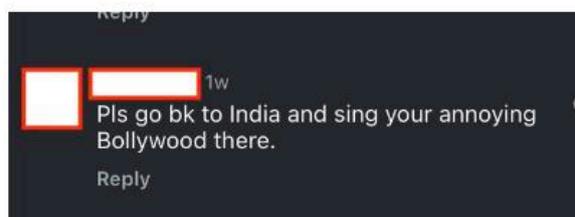


Figure 17 - Many comments were not explicitly inciting hate but could still be harmful.

Overly complex policy and governance documents

Governance documents and transparency reports remain lengthy, highly technical, and overall difficult to navigate. Although Instagram's terms and policies are available to users, their level of complexity makes them unlikely to be understood without the guidance of an adult or an expert. For instance, Meta's Platform Policy is over 6500 words long and has a Fleisch Reading Ease Score of 48.4 (putting it at a 'college' reading level).

Conclusion

Our testing of Instagram's Teen Accounts suggests that, in practice, key protections for young users often fall short of meaningful impact. The avatar accounts we created were still exposed to obscured advertising, inappropriate content, toxic conduct and addictive design. The governance measures in place during the testing period did not meaningfully mitigate these risks.

While acknowledging that Teen Accounts are a step forward, the changes we observed appeared limited in their effect without the adoption of privacy- and safety-by-design principles. Similarly, post-design measures such as moderation did not provide the level of protection needed to shield teen users from avoidable harms.

The preliminary findings raise significant legal compliance concerns – notably relating to the UK's Online Safety Act and the ICO's Age Appropriate Design Code (AADC) as well as the EU's Digital Services Act, AI Act and General Data Protection Regulation.

