Joint Committee on Children, Equality, Disability, Integration and Youth
16 Aibreán 2024, 3pm
Opening Statement: Meta

Go raibh maith agat a Chathaoirligh,

I'd like to thank the members of the committee for the invitation to appear before you today to discuss the subject of the protection of children in the use of artificial intelligence. My name is
Dualta Ó Broin, I'm Head of Public Policy for Meta in Ireland. I'm joined by my colleague David Miles, who is Safety Policy Director for Europe, Middle East and Africa with Meta.

For the convenience of the committee members, I've attached as an appendix to this statement a recent submission which Meta made to the Joint Committee on Justice as some of the topics contained in it are relevant to today's discussion.

Introduction

While Meta believes in freedom of expression, we also want our platforms, Facebook and Instagram, to be safe places where people and in particular young people, don't have to see content meant to intimidate, exclude or silence them.

We take a comprehensive approach to achieving this; by writing clear policies, Community Standards[1] in the case of Facebook and Community Guidelines[2] in the case of Instagram about what is and isn't allowed on our platform; by developing sophisticated technology to detect and prevent abuse from happening in the first place; and by providing helpful tools and resources for people to control their experience or get help.

We regularly consult with experts, advocates and communities around the world to write our rules and we constantly reevaluate where we need to strengthen them.

---

[1] https://transparency.fb.com/en-gb/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards
[2] https://www.facebook.com/help/instagram/477434105621119/

Use of AI technology in removing harmful content

AI plays a central role in reducing the volume of harmful online content on Facebook and Instagram.[3] Our online and publicly accessible [transparency centre][4] contains quarterly reports on how we are faring in addressing harmful content on our platforms, in addition to a range of other data. This includes how much content we remove, across a broad range of violations, and how much of that content was removed before any user reported it to us.

There are some violation areas where AI is extremely effective - for example fake accounts where over 99% of violations are identified by our AI systems.[5]

An example of a more difficult violation area for AI is bullying and harassment. In this area we removed 7.7 million posts from Facebook and 8.8 million posts from Instagram in OctoberDecember 2023. 86.5% on Facebook and 95.3% on Instagram were identified by our AI systems and removed before they were reported to us by a user. One of the reasons that AI is not as effective in this harm area yet is that bullying and harassment can be quite contextual - and not as immediately apparent as a fake account.

These systems are constantly improving, the same metric for the bullying and harassment violation for October-December 2022 was 61% in the case of Facebook and 85.4% in the case of Instagram.[6]

Tools for parents and teens

In addition to the actions we take to remove harmful content, we have built over 30 tools and features that help teens have safe, positive experiences and give parents simple ways to set boundaries for their teens. I have included a link to the timeline of these tools [here.][3] Further information about these tools and features, and how they work, can be found in our Instagram

---

[3] Our content enforcement relies on a combination of people reporting content, AI technology and human reviews.

[4] https://transparency.fb.com/

[5] https://transparency.fb.com/reports/community-standards-enforcement/fake-accounts/facebook/

[6] 6.4 million and 5 million pieces of content removed respectively.
https://transparency.fb.com/reports/community-standards-enforcement/bullying-andharassment/facebook/

---

[3] https://www.meta.com/en-gb/help/policies/safety/tools-support-teens-parents/

Parent Guide[4] our Family Centre[56], and additional resources on supportive online experiences can be found in our Education Hub[10] for Parents and Guardians.

Age verification

While these centres and guides give parents the ability and resources to navigate our tools and products, we understand that it can be overwhelming for parents to stay on top of every new feature and product across every application. In the US the average teenager uses 44 applications on their phones.

We believe that a significant step forward can be taken at a European level to ensure that parents only need to verify the age of their child once - and that their child will then be placed into an age appropriate experience on every single app. The most efficient and effective way in which this would work would be at the operating system or app store level. This would not remove responsibility from every app to have processes in place to manage age effectively and my colleague David can go into the steps that we at Meta take.

The question of age verification is complicated, however we believe that the time has come to move forward with an effective solution that addresses the concerns of all stakeholders including parents.

Education

I know that the importance of supporting the education system was mentioned in a previous session of this Committee. Meta is very proud to have provided the funding for the establishment of the FUSE programme by the National Antibullying Centre in DCU, and we welcome the announcement by Minister Foley last week about its further development[7]. The programme, which is available to every school in Ireland is currently being delivered in 440 primary and 570 post-primary schools.

---

[4] https://about.instagram.com/community/parents
[5] https://familycenter.meta.com/
[6] https://familycenter.meta.com/education/
[7] https://www.gov.ie/en/press-release/e831f-minister-foley-announces-new-anti-bullying-initiatives-forschools/

FUSE is the world's first research-based Anti-Bullying and Online Safety Programme designed to comply with UNESCO's Whole Education Approach to tackling bullying and online issues in schools.

We are pleased that in the past the Department of Justice has added modules to the programme and that the Department of Education is now supporting it. There may be scope in evolving existing modules or developing further modules to adapt and respond to evolving technology such as AI.

Submission to the Joint Committee on Justice

As set out in our submission to the Justice Committee in March, as part of Meta's commitment to transparency, we have [published more than 20 AI system cards](#) that explain how Artificial Intelligence powers recommendation experiences on Facebook and Instagram. In that submission, we described the way in which we use these systems to improve the user experience and make it safer and we also described the tools and controls available to users to control their experiences.

Conclusion

Finally, it is sometimes claimed that Meta is financially motivated to promote harmful or hateful content on our platforms to increase engagement. This is simply untrue. This content violates our policies and prevents our users from having enjoyable experiences. As a company, the vast majority of our revenue comes from advertising. Our advertisers do not want to see such content next to their ads. It is clear therefore that we are financially motivated to remove content which violates our policies as quickly as possible once we become aware of it.

I hope that this gives members of the committee an overview of some of the uses of AI by Meta, and we look forward to your questions.