yubo

afnor GROUP

AFNOR

SPEC 2305

Risk prevention and protection of minors on social media platforms

Executive summary

Online safety is a complex issue so, if you are a new social media platform, where do you start? These guidelines have been created to share knowledge and encourage best practice. They aim to help platform providers verify who is using their service, put detection, moderation and reporting measures in place, and build awareness and trust.

The result of a collaboration between international social media, online safety and child protection experts (e.g. Meta, Yubo, Dailymotion, Yoti, Bodyguard, E-enfance...), the guidelines focus on French and European regulations and legislation. But we hope they will be adapted in line with regulatory obligations and local laws in other countries so that even more young people around the world can get the most out of their digital spaces.

Account verification

As social media continues to increase in popularity, platform providers must understand the risks related to malicious accounts and take action to verify their users without compromising the user’s rights and privacy. The term ‘malicious account’ means any account created on a social media platform with the aim of causing harm to an individual, undermining the safety of the platform, or for fraudulent purposes.

We recommend that platform providers map the risks for their specific platform and audience (e.g. disinformation, cyberbullying, online hate speech, identity theft and grooming) and keep a risk register that is updated regularly. To help them take action against malicious accounts, we provide advice on how to spot the warning signs and how to have a proportionate response that strikes the right balance between data minimisation, identified risks and legal obligations.

Measures aimed at detecting and banning malicious accounts involve the processing of personal data but it is important that the principles of Safety and Privacy by Design are observed in order to minimise the impact on the user’s privacy and to comply with legislation. Whilst providers should collect as little data as possible, there are some situations that require reliable identification data to be obtained.

In these guidelines, we provide detailed information about the three most common account verification measures:

Age checks

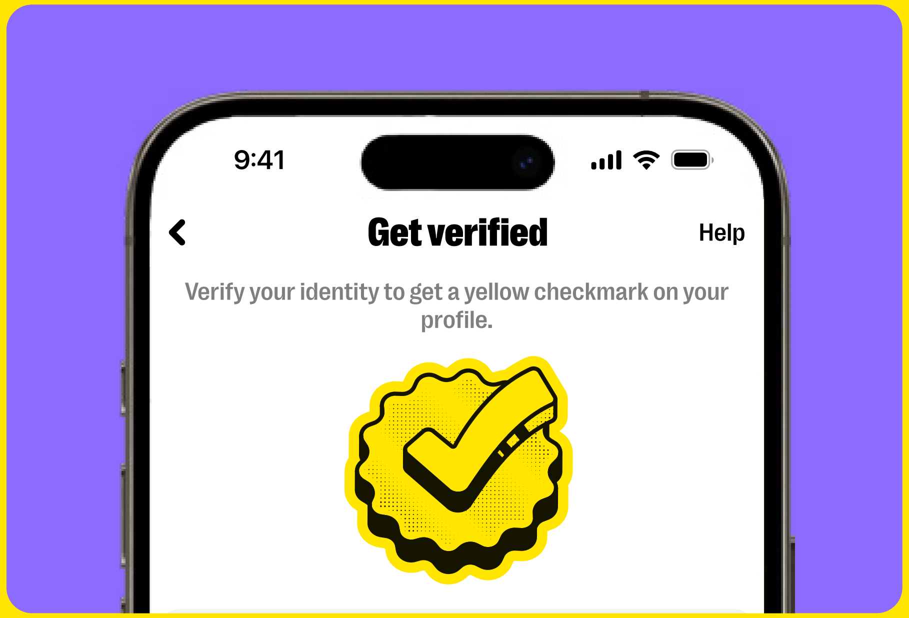
This includes ‘Age verification’(checking an individual’s exact age based on their date of birth – this usually requires the provision of an official identity document) and ‘Age estimation’ (making an approximate guess at an individual’s age – this is usually done by an artificial intelligence system that analyses an individual’s facial features on a photo).

Parental consent

In France, for example, minors aged under 15 cannot create a social media account without the authorisation of at least one adult with parental authority.

Identity verification

In some cases, platform providers ask their users to provide identity documents, such as a passport or driving licence.



Get verified

Verify your identity to get a yellow checkmark on your profile.

We also explain how any account verification measures must be made clear to users in the privacy policy, terms of use and other documents, and how platform providers should conduct a Data Protection Impact Assessment (DPIA).

Detection, content moderation and reporting

Platform providers face a major challenge – how to help users remain safe without undermining freedom of expression and ensure compliance with legal obligations.

Combating illegal online content, such as terrorism, incitement to discrimination and child sexual abuse content, must be a priority. Providers are obliged to set up effective means for removing any content reported as being “manifestly illegal” and to report any serious violations to the competent authorities. We provide advice on establishing standards, policies and community guidelines and ensuring that these are continually improved.

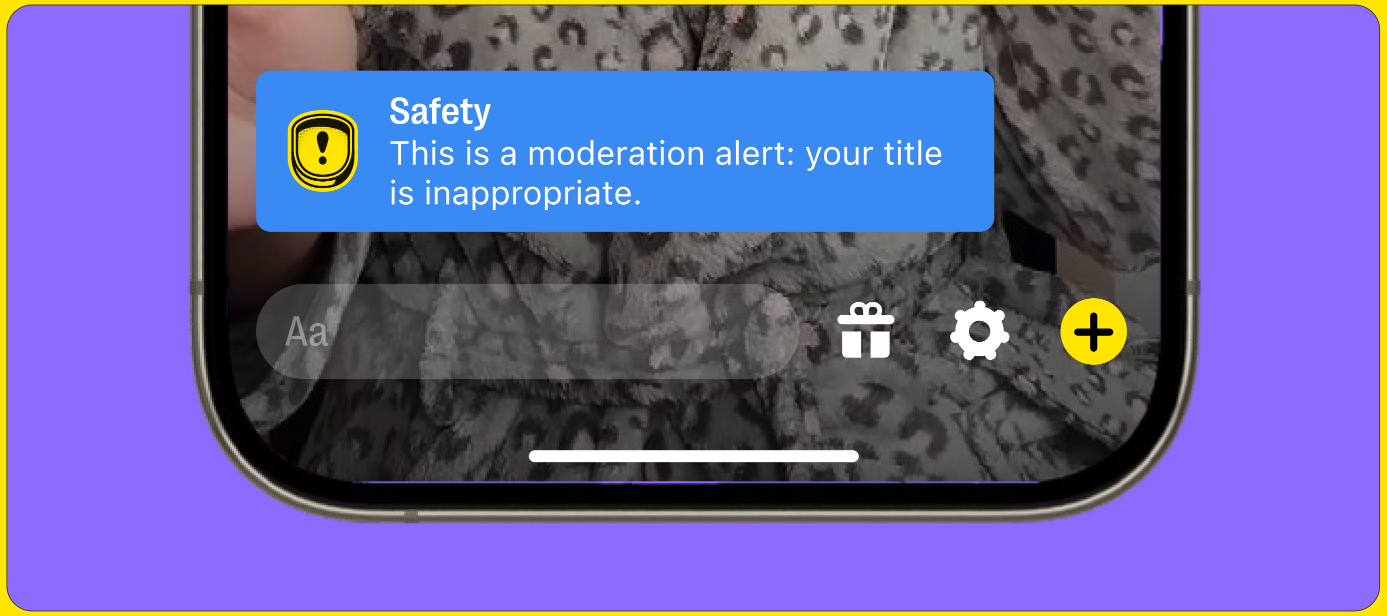
It is also essential that providers have content moderation tools for detecting inappropriate and/or illegal content. This includes proactive detection (detecting and analysing content as soon as it is submitted, either by a machine or a human being) and reactive detection (where users report the content to the platform for analysis and a moderation decision).

At a content level, moderation actions might include removing content, restricting content and sending cautions to users. Platform providers can also use dissuasive solutions (e.g. closing a user’s account temporarily) and positive incentives (e.g. giving a user a badge for respectful behaviour).

Due to the large volume of content on many platforms, content moderation consoles, automated tools and other technical resources should be leveraged. But human Trust & Safety professionals also play a vital role. We recommend making smart recruitment choices, providing ongoing skills development and, because of the nature of their work and the type of content they might be exposed to, prioritising the mental health of Trust & Safety team members.

Finally, platform providers have to address two main issues regarding cooperation with law enforcement and other authorities, such as PHAROS (France) and the National Center for Missing & Exploited Children (NCMEC) (US) – 1) Reporting illegal content and emergency situations involving a risk to human life and 2) Responding to legal requests.

To help with this, we provide guidance on topics such as correctly identifying content, storing data, communicating with the authorities, aggregating data and creating a continual improvement system.



Safety

This is a moderation alert: your title is inappropriate.

Transparency and raising awareness

Social media platforms should refer to the United Nations Convention on the Rights of the Child (CRC) and the General Data Protection Regulation (GDPR) and endeavour to create an environment of trust (both for under 18s and the adults responsible for them) as well as raising awareness of key online safety issues through partnerships.

To build transparency and trust, we recommend that platform providers:

- Create community guidelines that explain the code of conduct to users and promote safe and responsible behaviour on the platform – for minors, these should not be long and technical documents
- Publish specific information for parents, educators and legal guardians on their website
- Inform users about how their personal data is collected and processed – as well as having a privacy policy, providers should send relevant messages to users at key moments and make it easy for them to make choices about the data they share
- Allow users to exercise their rights under the GDPR, which include the right to access personal data held by the provider and the right to request the erasure of personal data

To help educate people and raise awareness of online safety issues, platform providers should collaborate with the government, NGOs, charities and other organisations. This might include:

- Developing campaigns for their own users (e.g. in-app campaigns)
- Supporting national and international awareness days, such as Safer Internet Day
- Attending conferences and other events
- Working with researchers and academics
- Creating tools for parents, educators and other adults (e.g. webinars, guides and videos)

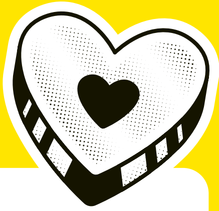
Legal and regulatory obligations

Several governments have adopted a proactive legislative policy to help protect minors online. So the AFNOR SPEC working group has included a non-exhaustive overview of the main legal and regulatory obligations of social media platforms in these guidelines.

The Legal Annex includes a summary of rights and fundamental freedoms (e.g. freedom of expression and principle of equality), personal data protection (e.g. GDPR and the e-Privacy Directive), criminal law (for offences such as grooming and sharing child sexual abuse material) and other relevant obligations.

It should be noted that these laws and regulations often change so the details in this section might be quickly out of date. Platform providers should, of course, investigate which obligations apply to them based on their own business model and local legislation.

Best practice



We hope these guidelines are a helpful resource for new platform providers that are committed to keeping young people safer on social media. By putting the right policies, tools, teams and partnerships in place and fulfilling their legal and regulatory obligations, they will be able to support the safety and wellbeing of minors without undermining their digital rights.

Thanks!