UK Council for Internet Safety

TRUSTED FLAGGER PROGRAMMES

GUIDELINES AND BEST PRACTICE

Many online platforms operate a 'Trusted Flagger' programme, recognising particular organisations in submitting reports of content, often on behalf of others to improve user experience and safety as well as trust.

Authored by a collaboration of organisations who operate Helplines that support victims of online harmful content and routinely report content to online platforms, this represents the principles and expectations for any Trusted Flagger programme.

It recognises and appreciates the online platforms who operate trusted flagger programmes and that each programme will differ based on a series of aspects, for example platform community standards, policies, resources, but that these principles represent the expectations from those 'Trusted Flaggers', as well as the obligations on them.

PRINCIPLES

- A general contact (eg email or system) for use by the trusted flagger - there must be a way of keeping this up to date and known by the trusted flagger, managing any turnover in staff at the platform for example.
- Nominated escalation point - if there is no response from the email address given to the trusted flagger, there must be another contact available for escalation.
- Clear communication about types of case which can be escalated in this way and what can't, with clear information about what to do with cases that can't be escalated in this way, advice, links to alternative routes etc.
- Where regulations allow, the trusted flagger can escalate on behalf of a user, recognising that Trusted Flaggers are often reporting on behalf of a third party.
- Regular meetings between the service provider and trusted flagger, for example twice a year, to share concerns, raise key issues and discuss progress, and review the functioning of the system
- Links to platform policies, specifically Terms and conditions, Community standards and Privacy policy. Trusted Flaggers will be automatically and routinely notified as and when these are updated

UK Council for Internet Safety

EXPECTATIONS

PLATFORM

- A recognition that any request for further information from the Trusted Flagger is limited, reasonable and realistic, for example Trusted Flaggers will not typically be able to provide proof of parental responsibility, proof of age, proof of identity
- That the platforms will conduct initial searches for relevant information to expedite the reporting process and before requesting of the Trusted Flagger.
- Expected response time; both automated and actioned and any operational changes to those should be agreed, for example within 24 hours.
- Platforms should share with trusted flaggers if there is any systemic delay, backlogs for example.
- Platforms should be ready to reopen a case if additional information comes to light and will review with any new context in mind.
- Trusted flaggers should be trusted, particularly being mindful of the additional context they can provide. Because of this, providers would be encouraged to adopt automatic suspension of content reported by a Trusted Flagger pending review
- Details of actions taken or reasons for rejection should be shared back with the trusted flagger.

TRUSTED FLAGGER

- To uphold the Trusted Flagger obligations and recognise that this status can be withdrawn by the Platform if misused.
- Use the Reported Flagger process as designed and for appropriate issues.
- Provide adequate and appropriate information to enable the platform to process the report, recognising that the more information provided will support a more efficient response
- Keeping Trusted Flagger reporting contacts confidential
- Uphold any confidentiality of victims