

Professor emerita Lorna Woods OBE

Violence Fixated Individuals
The Online Safety Act and Other Legislation

Introduction

- Overview of Part 3 OSA regime
- Questions about scope and implementation
- Image deletion orders and other orders

Overview of Online Safety Act Part 3 Regime

User-to-user • Social media • Messaging app • Gaming • Online marketplaces	Risk Assessment • Illegal content • Content harmful to children • Children's access assessment	Safety Duty • System for taking content down only in relation to illegal content • Age verification only required for primary priority content	Cross-cutting duties • Content reporting and complaints • FoX/privacy • Record keeping and review	Category 1 • Fraud • Public interest content • User empowerment • FoX/privacy • Terms of service • Transparency Category 2B • Transparency
Search • Includes ChatGPT	Risk Assessment • Illegal content • Content harmful to children • Children's access assessment	Safety Duty • No take down requirements • No express age verification requirement	Cross-cutting duties • Content reporting and complaints • Freedom of expression/privacy • Record keeping and review	Category 2A • Transparency

Key Points

- **Systems based approach; not aimed at particular items of content**
- **Precise obligations vary depending on**
 - Distinction between illegal content and content harmful to children
 - Nature of service – obligations on search service less far-reaching
- **Limited protection for adults**
- **Role of Ofcom**
 - Guidance, codes and “safe harbour”
 - Enforcement – note preference for cooperation and less formal supervision

Ofcom's Codes

- **Illegal content code focussed on governance and moderation**
 - See eg concessions from X around speed of dealing with terrorist content (15 May 2026)
- **Limited use of hash matching**
 - this is being extended, but still to limited areas of content
- **Limited controls on recommender systems for illegal content**
 - cf Children's Safety Code

Additional Measures

- Crisis response protocols
- Extended use of hash matching, bringing in NCII
- Automated tools to detect fraud, suicide content
- Changes to recommender tools requirements for illegal content
- Proactive technologies to detect grooming and CSAM

Who does OSA empower?

- **Not designed for crises or fast moving events**
- **Focus on content specific, ex post measures**
 - Might use of more content neutral interventions be helpful?
- **Speed of investigations**
 - No timescales in OSA;
 - What level of evidence is required – especially for business disruption measures?
 - Threat of litigation
- **Information sharing**
 - Trusted flaggers
- **Platforms still have significant say in what is permissible**

The Impact of Content Categories as Trigger for Duties

- **Reliance on content categories to trigger regime**
 - Determination of content not role of choice architecture
 - Emphasis on priority categories
 - NB potential significance in this context of eg s 127 Communications Act
- **Problems from “illegal content”**
 - online/offline distinction
 - Sub-criminal content – is it all about children and priority content?
 - Does age verification solve all problems for under 18s anyway?

Ofcom Codes

- **Narrow; disconnect between identified risks and required measures**
 - Technical feasibility
 - Proportionality
 - Clarity
- **individual item focussed “harm” allows for cumulative impact but hard to put this in illegal content framing**
- **Emphasis on ex post moderation**
 - Role of services in facilitation and incentivisation seemingly not considered
- **Search**
 - Is one click limit on content for which they are responsible appropriate?

Proactive Technology

- **Can be used in both codes and enforcement mechanisms**
BUT
- **Conditions on use, especially FoX and privacy protections**
 - Need for safeguards
 - Appropriate balance between precision and recall – consistency across services?
 - Remedies and appeals
 - Is this as simple for eg behavioural tools as for content tools?

Terms of Service

- Ofcom expects most moderation to be done by reference to service's own terms of service/community standards
- Services cannot take measures against content or user unless specified in terms of service (or because of safety duties)
- Category 1 is only group which must enforce terms of service
- No minimum threshold; no protection against roll back

Individuals not content

- **Criminal law attaches to (user) behaviour, not content; applicability of regime depends on individual users' mindset**
 - Mental element and impact on viral content
 - Is terrorism “exception” enough?
- **Not just a problem for OSA enforcement but for victims seeking redress**

Ofcom: ICJG para 1.56
For other priority offences,
however, the illegality of the content in each
new iteration will depend upon the likely state
of mind of the person sharing, forwarding or
reposting it.

Responses to NCII

- **Additional measures**
 - changes to illegal content code to require hash matching
- **Section 100 CPA amends OSA illegal content duties**
 - system to take down (U2U) or ensure users are no longer able to encounter (search) notified content and content substantially similar within 48 hours
- **Ofcom will consult on code changes by end of 2026**
- **Failure to comply with confirmation decision is offence**

Responses to NCII, cont

- **Why 48 hours?**
- **Still systems based so may not be effective in all cases -is this good enough for victims?**
 - NB OSA does not provide for an individual route to redress; victims must use services' complaints systems or legal routes to action
- **Only applies to NCII (not CSAM); not other priority illegal content, including terrorism**

Image Deletion

- **Section 102 (not yet commenced and prospective only)**
- **Image deletion order – based on criminal offences: intimate image offence, breastfeeding voyeurism recording, or sharing semen-defaced images**
- **“deletion” – differentiate between different media**
- **In line with existing norms?**
- **Viral content – back to OSA?**
 - Or civil law mechanisms?

In the case of intimate image offences [...]), devices which were used or intended to be used to commit or facilitate these offences are liable to a deprivation order and the prosecutor should remind the court of its powers in this respect. This is particularly important in respect of devices which retain the images which the offender shared or threatened to share. Note that this equally applies to images which appear to show a person, and therefore extends to deprivation orders for offending involving pseudo-images or deep fakes.

Current Guidance from CPS on Communications Offences

Gaps and Weaknesses NCII cont, image deletion

- **Who is in control of asking for orders?**
 - CPS historic performance under s 153 Sentencing Act weak
 - NB changes to Communications Offences guidance in 2025 prompting prosecutors to make better use of deprivation orders.
 - Cf previous practice in re CSAM – usual practice to make order
- **Court to give reasons for not making order**

Weapons Content

- **CPA introduces Content Removal Notices**
 - removal of specific content deemed to be illegally marketing or selling knives or offensive weapons under existing legislation (which is listed in Sch 7 OSA)
 - 48 hour time frame
 - Right of review
 - backed up by civil penalty

Commentary on Take Down

- **NCII and weapons content have a common 48 hour timeframe but are otherwise different in who triggers the take down and whether the obligations is systems-based or content-based**
 - Is this calibration to circumstances or just random?
- **Why these two types of content and not others?**
 - Are there other responses to lower level content that should be considered?
- **Other routes to take down (theoretically available for NCII) might not be available for other types of content**

Other Powers

- **Digital Economy Act 2010**
 - Concerns abuse of domain names as regards UK connected registries
- **CPA 2026, Sch 18**
 - UK law enforcement agencies have no legal power to require registrars and registries to suspend domains but rely on voluntary arrangements
 - Law enforcement agencies may apply to Court for a DNS order of up to 12 months' duration in respect of "serious crime"
 - Use of MLATs/international cooperation as fall-back.
- **NB potentially serious consequences for speech/collateral censorship**