



Risk & Responsibility

Online Safety, free speech & defamation

SSLC comments

- Positive comments about seminars (acknowledging attendance is not good)
Thanks, and yes, it's worth attending 😊
- Mixed opinions about whether the exam setup (without crib sheet) has been communicated well. Some thought more information would help. **I've been careful to make the exam scenarios and questions more general than last year's exam, so answers will be more about their ability to identify relevant issues, reflect on dilemmas and discuss the impact for different groups rather than knowing very detailed aspects of specific laws or professional codes**
- Positive comments about the slight adjustment in timing for the coursework.
Thanks
- Some comments about slightly adjust the ordering of lecture materials as some of the legal topics appear later. **It's difficult in a 10-week course to cover everything that is needed before the coursework is due in. I believe I will have covered the major topics by then.**

Legislating for online experience

- Rise of harmful and illegal content online has led to concern
- Through social media, users are now interacting a lot more with each other, which opens them to harmful encounters such as trolling, abuse and the sharing of offensive or illegal material
- There are concerns that children and young people are also experiencing this type of harm, putting many in danger of being exposed to illegal and harmful content
- On the other hand, many believe that free speech is an important tenet to protect in online spaces ...

Online Safety Act 2023

- Aims to protect children and adults online
- Creates a set of duties for social media companies and search services, making them more responsible for their users' safety on their platforms
- Gives providers duties to implement systems and processes to reduce risks their services are used for illegal activity, and to take down illegal content
- The strongest protections are designed for children. Platforms will be required to prevent children from accessing harmful and age-inappropriate content and provide parents and children with clear and accessible ways to report problems online when they do arise.
- Also protects adult users, requiring major platforms to be transparent about which kinds of potentially harmful content they allow, and give people control over the types of content they want to see.

Criminal offences in OSA 2023

- The criminal offences introduced by the Act came into effect on 31 January 2024. These offences cover:
 - sending false information intended to cause non-trivial harm
 - threatening communications
 - encouraging or assisting serious self-harm
 - cyberflashing
 - intimate image abuse

Protections in OSA 2023

- Requires online services to:
- Take robust action against against illegal content and activity i.e.:
 - Child sexual abuse, Controlling or coercive behaviour, Extreme sexual violence, Fraud, Racially or religiously aggravated public order offences ...
- Take robust action against content that is harmful to children, i.e.:
 - Pornography, content that promotes self harm, eating disorders, suicide ...
- To enforce age limits consistently and protect child users
- To offer adult users tools to give them control over content they see
- To rapidly remove suicide and self-harm content



- Molly Russell, 14, took her own life in November 2017 after viewing extensive content, particularly on Instagram and Pinterest, related to suicide, depression, self-harm and anxiety
- The inquest concluded that social media contributed to her death, stating that she “died from an act of self-harm while suffering from depression and the negative effects of online content”
- Inquest comments fed into decisions about OSA provisions

OSA Regulator

- OFCOM is the independent regulator for the OSA www.ofcom.org.uk
- It sets out steps providers can take to fulfil their safety duties in codes of practice and advice.
- It has a broad range of powers to assess and enforce providers' compliance with the framework – from running audits to fining companies
- Providers' safety duties are proportionate to factors including the risk of harm to individuals, and the size and capacity of each provider.
- This makes sure that while safety measures need to be put in place across the board, small services with limited functionality don't need to take the same actions as large corporations.
- Ofcom is required to take users' rights into account when setting out steps to take. And providers have duties to take regard of users' rights when fulfilling their safety duties.

Current codes of practice for OSA

- Companies need to work with Ofcom to comply
- Providers of online user-to-user services will be asked to
 - Take proportional steps to prevent users encountering illegal content (i.e. perpetrating fraud, sharing child sexual abuse material)
 - Manage the risk of offences taking place through the service
 - Be able to do age verification if necessary
 - Undertake a risk assessment and manage risks identified
 - Remove illegal content when aware of it
 - Protect children from harmful content
 - Explain the approach taken in terms of service
- More stringent measures for large services, less for smaller ones

Benefits of the OSA

- **Child protection** - enforces age-checking measures to prevent children from accessing age-inappropriate content on social media platforms.
- **Content removal** - requires tech companies to promptly remove illegal content like child sexual abuse material, hate speech, and content promoting self-harm.
- **User control** - allows users to filter out unwanted content like online abuse and customize their online experience.
- **Transparency and risk assessment** - forces large platforms to publish risk assessments outlining potential dangers to users, particularly children.
- **Accountability for platforms** - holds tech companies responsible for enforcing their own terms of service and taking action against harmful content.
- **Improved reporting mechanisms** - provides clear pathways for users to report harmful content and concerns.
- **Enforcement by Ofcom** - grants the communications regulator Ofcom with power to investigate and sanction companies that fail to comply with online safety rules.

Criticisms of the OSA

- **Over-reach on content moderation** - could force platforms to remove too much content, potentially impacting legitimate discussions and viewpoints, due to broad definition of "harmful content" / not removing enough content, or doing it quickly enough
- **Privacy concerns** - requirement for platforms to scan private messages to detect illegal content with fears of government surveillance
- **Free speech implications** - could stifle free speech by causing platforms to err on the side of caution and remove content that may be controversial but not necessarily illegal
- **Enforcement challenges** - how effectively the Act can be enforced, particularly with regards to identifying and removing harmful content across a wide range of platforms
- **Potential for abuse** - could be used to target specific groups or viewpoints, leading to censorship
- **Lack of focus on system design** - focuses on content removal and not enough on proactive measures to prevent harmful content from appearing in the first place
- **Unclear definitions** - some provisions in the Act lack clear definitions, which could lead to inconsistent application and interpretation

Online Safety Act not fit for purpose after far-right riots, says Sadiq Khan

Exclusive: London mayor warns law must be revisited 'very, very quickly' due to falsehoods that contributed to unrest

● [Sadiq Khan on the riots: 'Like a lot of people of my generation, I felt triggered'](#)



📷 Riot police facing disturbances outside a Holiday Inn Express in Rotherham on Sunday. Photograph: Christopher Furlong/Getty Images

Laws designed to counter misinformation are “not fit for purpose” and must be revisited after the spread of online falsehoods contributed to this month’s far-right riots, the mayor of London has said.

Sadiq Khan, one of the UK’s most senior Muslim politicians, said ministers should act “very, very quickly” to review the Online Safety Act after the violent unrest in England and Belfast over the past week. There have been calls to hasten the act’s implementation.

His comments came after the owner of X, Elon Musk, escalated his attacks on the Labour government, [sharing a fake Telegraph article](#) on his social media platform claiming Keir Starmer was considering sending far-right rioters to “emergency detainment camps” in the Falklands. The article was first posted by Ashlea Simon, a co-leader of the far-right group Britain First.

THE ONLINE SAFETY ACT

BIG BROTHER WATCH TEAM / SEPTEMBER 16, 2025



2. Social media platforms have formally become judge and jury over our speech

At the heart of the Online Safety Act is the delegation of responsibility for individuals’ online expression to social media platforms themselves. This approach runs contrary to the general principle that people should ultimately be responsible for their own actions and in the context of a communications network, only incentivises these platforms to censor where they are threatened with penalties if they do not. Whilst some illegal content will always be clear and obvious to content moderators, it is inconceivable that they should be able to make determinations on what might legally constitute “stirring up hatred” or “malicious communication”, speech which can reach a criminal threshold but which the police and the courts frequently find hard to make judgments on. Silicon Valley’s content moderators can’t possibly fulfil the tasks of police, judge and jury, so when these difficult determinations are presented to them, under the threat of penalties, they will almost certainly censor lawful speech out of an abundance of caution.

Wikimedia Foundation v SoS SIT 2025

Wikipedia argued that OSA could undermine its model of anonymous, collaborative editing if classified as Category 1 service

Wikipedia's challenge

- ***Risk of mandatory verification***: could compel it to verify the identities of its editors
- ***Threat to its model***: either lose contributor anonymity or implement unworkable rules.
- ***Vandalism and misinformation***: OSA provisions allowing users to filter out or block non-verified users could allow malicious actors to easily remove content.
- ***Incompatible with purpose***: the purpose of an online encyclopedia is fundamentally different from social media sites, so applying the same rules doesn't make sense.

Outcome of the challenge

- Dismissed by the High Court, ruling that the Secretary of State acted lawfully in setting the regulations
- However, it did not give the government a "green light" to impose obligations that could damage Wikipedia's operations
- The judgment emphasised the duty of regulators (Ofcom and the UK government) to ensure that platforms like Wikipedia are protected from being unfairly harmed

Freedom of expression

FoE is protected by Article 10 of the UK Human Rights Act 1998:

“Everyone has the right to freedom of expression ... to hold opinions and to receive and impart information and ideas without interference by public authority ...”

Restrictions include:

- **Protecting others:** rights and reputations of others or preventing disorder or crime.
- **Preventing hate speech:** including incitement to racial or religious hatred, or terrorism.
- **Protecting confidential information:** disclosure of information received in confidence.
- **Maintaining judicial impartiality:** protecting the authority and impartiality of the judiciary.
- **Religious freedom:** can criticise religion but not incite religious hatred
- **Anti-protest legislation:** if the protest may result in public disorder, property damage, or disruption to the life of the community, or in support of a proscribed terrorist group.

Defamation

- Defamation is a statement injurious to one's reputation, e.g. by facilitating 'hatred, ridicule or contempt' or lowering one's standing in society's view
- Why is protecting reputation important?
 - Reputation is a civil right of the same standing as any other, e.g. right to enjoyment of life
- Covered by Defamation Act 2013
- Defamation is a civil wrong, or tort. Claims generally have to be issued in the High Court, only sometimes in the County Court

Libel and Slander

- In UK law, libel and slander are types of defamation, which protect a person or organisation's reputation from false statements:
- **Libel**
 - Defamatory statements that are written or broadcast, such as in articles, blogs, social media posts, or other written content.
- **Slander**
 - Defamatory statements that are not published and are transient, i.e. conversation, impressions, other transient media that are not stored

What is defamatory

- Three tests:
 - The statement is potentially and actually defamatory
 - The statement is clearly about the claimant
 - The statement was published/communicated
- **Serious harm:** Claimants must prove that the publication of a statement caused or is likely to cause serious harm to their reputation. For companies trading for profit, the harm must also cause or be likely to cause serious financial loss. Can't defame a government body.

Defence against defamation

- **Defences:** The Act outlines defences for defamation:
 - **Truth:** A statement can be fully defended against defamation if it is substantially true.
 - **Honest opinion:** A statement can be defended if it is an opinion based on accurate, referenced facts.
 - **Public interest:** A statement can be protected if it is on a matter of public interest and the defendant reasonably believed publication was necessary
 - **Operators of websites:** if complied with procedures and remove posts when a complaint is received
 - **Peer-reviewed statements in academic papers**
 - **Reports protected by privilege:** i.e. court reports

Wagatha Christie case

- **Rooney's viral post:** In **October 2019**, Coleen Rooney posted on social media accusing Rebekah Vardy's Instagram account of leaking fake, private stories to The Sun newspaper.
- **Vardy's lawsuit:** Rebekah Vardy, who denied leaking the stories, launched libel proceedings against Rooney in June 2020. The High Court later confirmed that Rooney's post was defamatory, but the trial would focus on whether Rooney's accusation was substantially true.
- **The 2022 high court trial and verdict**
- **Witness credibility:** The judge found Rooney to be an "honest and reliable witness". Conversely, she concluded that Vardy's evidence was "manifestly inconsistent with the contemporaneous documentary evidence, evasive or implausible".
- **Verdict:** Rooney proved her case: In a July 2022 ruling, the judge sided with Rooney, finding that her public post was "substantially true." The court found that Vardy "knew of and condoned" her agent's leaking of stories to The Sun.
- **Final settlement:** In **May 2025**, Vardy was ordered to pay over £1.4 million of Rooney's legal costs, bringing an end to the protracted legal battle.