Whose Fault Is It Anyway?
Section 230 & Liability for Online Harms

Tony Phillips | Partner, Compliance, Investigations & Complex Disputes | Washington, DC
Steven Farmer | Partner, Technology Transactions | London
Mark Booth | Associate, Technology Transactions | London
Johnna Purcell | Associate, Government Law & Strategies | Washington, DC
Alexis Wansac | Associate, Litigation | Washington, DC

Agenda

- **Section 230 Past & Present:** The Creation & Development of the Internet as We Know It
- **The Internet in Europe:** The E-Commerce Directive, the Digital Services Act, & the Online Safety Act
- **The Future:** The Impact of Generative AI on U.S. Regulation of Online Services
- **Conclusion & Q&A**

Section 230 Past & Present:
The Creation & Development of the Internet as We Know It

The Mind-Boggling Scale of the Internet

pillsbury

- **5.47 billion** unique internet users in the world
- **5 billion** internet searches per day
- **4.1 million** YouTube videos watched per minute
- **3.7 billion** daily active users of Meta's core products (Facebook, Instagram, WhatsApp, and Messenger)
- **1 million** Tinder swipes per minute
- More than **1.98 billion** websites online (and growing)

The 26 Words that Created the Internet

pillsbury

- "No provider or user of an **interactive computer service** shall be treated as the publisher or speaker of any information provided by **another information content provider**."
  - Section 230(c)(1) of the Communications Act of 1934 (enacted in 1996)
  - Goal of the legislation was to protect nascent online service providers from defamation claims arising from user-generated content
- Section 230 differentiates between those who create content and those who provide access to the content—protects the latter from liability for the former

Content Moderation Is Born

pillsbury

- **Section 230(c)(2):** No provider or user of an interactive computer service shall be held liable for good faith voluntary action "to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."
  - "Providers" of interactive computer service = Facebook, Snapchat, TikTok, Google
  - "Users" of interactive computer service = group moderators, page hosts, individuals
- Protects those attempting to moderate potentially harmful content from liability for missing something

Statutory Exemptions from Section 230 Protections

Five categories of claims to which Sec. 230 immunity provisions do not protect entities from liability:

1. Federal criminal prosecutions
2. Intellectual property disputes; however, may be protected under the Digital Millenium Copyright Act
3. State law claims that are not otherwise preempted by Section 230
4. Suits brought under the Electronic Communications Privacy Act and similar state law provisions
5. State criminal or federal civil violations of the Fight Online Sex Trafficking Act (i.e., COPPA, SESTA-FOSTA)

Judicial Limits to Section 230 Protection

pillsbury

Courts found that Sec. 230 protection didn't apply when:

- Defendants induced or contributed to the development of unlawful content.
  - *Fair Housing Council of San Fernando Valley v. Roommates.com*, 521 F.3d 1157 (9th Cir. 2008)
  - *FTC v. Accusearch*, 570 F.3d 1187 (10th Cir. 2009)
- Defendant failed to warn about illegal conduct of which it had actual knowledge.
  - *Doe v. Internet Brands*, 824 F.3d 846 (9th Cir. 2016)
- Defendant selectively reposts—and thus adopts—third-party content.
  - *Diamond Ranch Academy v. Filer*, 117 F.Supp.3d 1313 (D. Utah 2016)
- Defendants failed to act in good faith when moderating content.
  - *E-Ventures Worldwide v. Google*, 188 F.Supp.3d 1265 (M.D. Florida 2016)
  - *Enigma Software Group v. Malwarebytes*, 946 F.3d 1040 (9th Cir. 2019)

Emerging Trends in Section 230 Interpretations

pillsbury

- Frequently distinguishing between "interactive computer service provider" and "information content provider" to apply Sec. 230 protection
- Courts extend protection against liability arising from user content but not necessarily to website features that develop or solicit that content
- Recent examples:
  - Negligent design claims regarding features developed by providers are not exempt from liability. See *Lemmon v. Snap, Inc.*, 995 F.3d 1085 (9th Cir. 2021)
  - Automated content moderation tools integrated into algorithms creates information content not protected under Sec. 230. See *Dangaard v. Instagram*, LLC, No. C 22-01101 WHA, 2022 WL 17342198, at *4 (N.D. Cal. Nov. 30, 2022; *Liapes v. Facebook, Inc.*, No. A164880, 2022 WL 20680402 (Cal. Ct. App. Sept. 21, 2022)

pillsbury

The Internet in Europe:
The E-Commerce Directive, the Digital Services Act, & the Online Safety Act

The E-Commerce Directive

pillsbury

- Directive 2000/31/EC of 8 June 2000 (E-Commerce Directive)
- Included provisions relating to intermediary liability to tackle "*existing and emerging disparities in Member States' legislation and case-law*"
- Safe Harbors for different service providers: (i) mere conduits; (ii) caching; and (iii) hosting
- Article 15 goes on to prohibit states from imposing general monitoring obligation:

  "Member States shall not impose a general obligation on providers, when providing [conduit, caching, or hosting services], to monitor the information which they transmit or store, nor a general obligation actively to seek facts or circumstances indicating illegal activity."

The EU & the Digital Services Act

pillsbury

- Regulation (EU) 2022/2065 on a Single Market for Digital Services ("DSA") amends the E-Commerce Directive and seeks to "rebalance" the responsibilities of users, platforms and public authorities
- Applies to "intermediary service" providers: (i) mere conduits; (ii) caching providers; and (iii) hosting providers
- Contains comparable safe harbors to the E-Commerce Directive and maintains the general prohibition on monitoring—Article 8 DSA:

  "No general obligation to monitor the information which providers of intermediary services transmit or store, nor actively to seek facts or circumstances indicating illegal activity shall be imposed on those providers."

What's Changed under the DSA?

The DSA codifies the commonly accepted "Good Samaritan" principle, where providers do not leave the liability safe harbors by undertaking proactive measures in relation to illegal content.

What's Changed under the DSA?

pillsbury

The DSA includes new obligations on intermediary service providers, such as:

- Informing users in terms and conditions on restrictions imposed on the use of their service in relation to information provided by users, as well as policies, procedures, measures, and tools used for the purpose of content moderation
- Acting in a diligent, objective, and proportionate manner in applying and enforcing the restrictions referred to above
- Provide reports at least annually on any content moderation that they engaged in during the relevant period (including any orders received from authorities to take down illegal content);
- Put easy to access and user friendly "notice and takedown" mechanisms in place for all illegal content (with the ability to designate "trusted flaggers")
- Notifying law enforcement if it becomes aware (or suspicious) of a criminal offence involving a threat to life or safety or a person
- Effective internal complaint handling mechanisms
- Prohibits so called "dark patterns"

The Role of "Big Tech"

- Additional obligations placed on Very Large Online Platforms ("VLOPs") and Very Large Online Search Engines ("VLOSEs"), including:
  - Undertaking diligent assessments of systemic risks stemming from their systems or the uses made of their services (taking into account their algorithmic and content moderation systems)
  - Putting in place reasonable, proportionate, and effective mitigation measures tailored to identified risks which may include adapting content moderation processes

DSA – Key Points to Note

- Entered into force on 16 November and generally applicable from 17 February, 2024
- Some provisions applicable earlier—e.g., VLOPs and VLOSEs have obligations four months after being designated
- Fines can be up to 6% of total worldwide annual turnover

The UK Post-Brexit

- The safe harbors in the E-Commerce Directive were implemented in the UK pre-Brexit, but Article 15 no longer applies to the UK
- Intermediaries are being seen less as neutral conveyers of content, i.e. "Big Tech," and people are more aware of the harms that can be caused by online content, e.g., vaccine disinformation during the Covid-19 pandemic
- "As an independent country, the UK has the opportunity to set the global standard for a risk-based, proportionate regulatory framework that protects citizens online and upholds their right to freedom of expression."

The UK & the Online Safety Act

- The **Online Safety Act 2023** ("OSA") applies to "user-to-user" services with "links" to the UK
- Allocates duties of care proportional to the risks posed: (i) general duties on all services; (ii) additional duties for Category 1, 2A, and 2B services; and (iii) additional duties for services visited by children
- General duties include:
  - Undertaking a risk assessment
  - Taking or using proportionate measures to prevent users from encountering priority illegal content
  - Enabling content reporting
  - Operating proportionate systems and processes to minimize the length of time that priority illegal content is present and that illegal content is swiftly taken down on notice

What Will This Look Like in Practice?

pillsbury

- The OSA regulates systems and processes but not content or results—liability is therefore not strictly based on what content is available
- "Proportionate" measures to prevent priority illegal content will vary on the service—no obligation to use artificial intelligence tools, but terms of service must give information about any "proactive technology" that is used
- Proactive technology means:
  - (i) content identification technology;
  - (ii) user profiling technology; or
  - (iii) behavior identification technology

OSA—Key Points to Note

- Passed into law on 30 October, 2023
- Much of the finer detail will be included in codes of practice to be issued by OFCOM—these will also set out recommended compliance steps that meet the online safety objectives
- Penalties for non-compliance can be up to £18m or 10% of worldwide group revenue for the previous year

The Future:
The Impact of Generative AI & Mitigating Digital Risk

The Role of Artificial Intelligence

pillsbury

- AI could revolutionize content moderation—clearly being considered in the new laws (e.g., "proactive technology" under the OSA)
- AI is vulnerable to legal challenge and presents compliance concerns:
  - Lack of transparency / explainability
  - Perpetuating bias
  - Potential inability to understand context
- Distinctions premised on "creating" content or "moderating" content break down as applied to generative AI

Industry Warnings About AI

- Microsoft White Paper: **"Governing AI: A Blueprint for the Future"**
  - "We are the first generation in the history of humanity to create machines that can make decisions that previously could only be made by people."
  - Five-Point Blueprint for Governing AI:
    1. Implement and build upon new government-led AI safety frameworks
    2. Require effective safety brakes for AI systems that control critical infrastructure
    3. Develop a broader legal and regulatory framework based on the technology architecture for AI
    4. Promote transparency and ensure academic and public access to AI
    5. Pursue new public-private partnerships … to address the inevitable societal challenges
- Sam Altman testimony: "[I]t is vital that AI companies adhere to an appropriate set of safety requirements … [in] a governance regime flexible enough to adapt to new technical developments."

The Global Reaction

pillsbury

- Judicial bodies now beginning to consider the legal implications of AI
- Legislative bodies around the globe are grappling with how to regulate AI, e.g., the EU's **AI Act**
- China **Generative AI Regulation**—focus on accuracy and fake or harmful misinformation
- Brazil draft AI law (May 2023)—human rights-oriented, risk-based and governance-focused

U.S. Executive Branch Response

- White House Office of Science and Technology Policy white paper: "Blueprint for an AI Bill of Rights"—Five Principles:
    1. Safe and Effective Systems
    2. Algorithmic Discrimination Protections
    3. Data Privacy
    4. Notice and Explanation
    5. Human Alternatives, Consideration, and Fallback
- White House Voluntary Commitments – Seven leading AI companies (Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI) commit to build "Safe, Secure, and Trustworthy" AI technologies
- Biden EO 14110: **"Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"**—intended to harness the benefits of AI while mitigating risks to consumers and keeping people safe

U.S. Legislative Branch Efforts

pillsbury

- Senate Majority Leader Schumer (D-NY) announced his **SAFE Innovation Framework** in June
  - **Security:** safeguard national security; ensure economic security by responding to job loss
  - **Accountability:** address copyright, protect intellectual property, address liability
  - **Foundations:** align AI with our democratic values
  - **Explain:** ensure government and the public have needed information to foster trust
  - **Innovation:** support U.S. leadership in unlocking the potential of AI
- **Blumenthal-Hawley Framework:** establishes guardrails for artificial intelligence with specific principles to inform legislative efforts:
  - Promote transparency with notice to users and database of incidents
  - Protect consumers and kids with controls and restrictions
  - Establish an independent oversight body responsible for licensing and registration
  - Legal accountability for entities and through private rights of action
  - Defend national security and international competition through sanctions and export controls

Compliance Program Considerations

pillsbury

- To what extent will courts be more likely to view activities as content creation as opposed to content moderation or hosting?
- Is AI being used in compliance with best practices and emerging global standards?
- Are the methods being used safe, secure, and trustworthy?