Problematizing and Reconceptualizing Online Harm

Dr Alicia Cork
Prof Adam Joinson
Prof Danae Stanton Fraser
Dr Laura Smith
Dr David Ellis
What harms are included in this work?

Child sexual exploitation and abuse
Terrorist content and activity
Organised immigration crime
Modern slavery
Extreme pornography
Revenge pornography
Harassment and cyberstalking
Hate crime
Sale of illegal goods or services
Content illegally uploaded from prison

Sexting of indecent images by under 18s
Encouraging or assisting suicide
Incitement of violence
Extremist content and activity
Cyberbullying and trolling
Coercive behaviour
Intimidation
Disinformation
Violent content
Promotion of Female Genital Mutilation (FGM)
Advocacy of self-harm

Children accessing pornography
Children accessing in appropriate material (including under 13s using social media and under 18s using data apps; excessive screen time)

Harms in Online Safety Bill

bath.ac.uk
Other harms included in this work

- Fraud
- Scams
- Disinformation
- Identity theft
- Surveillance
- Microtargeting
- Profiling
- Data theft
- Addictions/excessive use
- Deepfakes
- Malware
- Hacking
- DDoS attacks
- Algorithmic bias
- Human trafficking coordination
- Excessive online gaming disorders
- Inaccessibility
- Misinformation
- Automated decision making
- Excessive online gambling disorders
- Threats
- Social engineering
- Destruction of data
- Technological misuse
- Blackmail
- Epilepsy trolling
- Unauthorized access to IoT devices
- Sexual harassment
- Pro-eating disorder content
Where we started

Figure 1: A Typology of Harmful Content
<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Contact</th>
<th>Conduct</th>
<th>Contract</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Child as recipient</strong></td>
<td><strong>Child as participant</strong></td>
<td><strong>Child as actor</strong></td>
<td><strong>Child as consumer</strong></td>
</tr>
<tr>
<td><strong>Aggressive</strong></td>
<td>Violent, gory, graphic, racist, hateful and extremist content</td>
<td>Harassment, stalking, hateful behaviour, unwanted surveillance</td>
<td>Bullying, hateful or hostile peer activity e.g. trolling, exclusion, shaming</td>
<td>Identity theft, fraud, phishing, scams, gambling, blackmail, security risks</td>
</tr>
<tr>
<td><strong>Sexual</strong></td>
<td>Pornography (legal and illegal), sexualization of culture, body image norms</td>
<td>Sexual harassment, sexual grooming, generation and sharing of child sexual abuse material</td>
<td>Sexual harassment, non-consensual sexual messages, sexual pressures</td>
<td>Sextortion, trafficking for purposes of sexual exploitation, streaming child sexual abuse</td>
</tr>
<tr>
<td><strong>Values</strong></td>
<td>Age-inappropriate user-generated or marketing content, mis/disinformation</td>
<td>Ideological persuasion, radicalization and extremist recruitment</td>
<td>Potentially harmful user communities e.g. self-harm, anti-vaccine, peer pressures</td>
<td>Information filtering, profiling bias, polarisation, persuasive design</td>
</tr>
<tr>
<td><strong>Cross-cutting</strong></td>
<td>Privacy and data protection abuses, physical and mental health risks, forms of discrimination</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Towards a Theory of Harms (Clark & Claffy, 2019):
- Harms to availability of internet access
- Harms to the integrity of the internet experience
- Harms to confidentiality and privacy
- Harms to innovation competition and choice
- Harms to journalism, the marketplace of ideas and the political processes that depend on them.

bath.ac.uk
Realisations:

- Linguistic inconsistency in what is meant by ‘Online Harm’
  - Need to discriminate between a risk and the harmful impacts of that risk
- Should we be looking to mitigate exposure to risks or actual harm experienced?
- Many different dimensions of online risk and harm worth considering:
  - Legality
  - Intentionality
  - Likelihood of exposure
  - Adult victims or child victims
  - Severity of harm
Systematic Review

- Systematic Review of all papers that contain a **Taxonomy of Online Harm or Risk**

- **Aims:**
  - Understand the **dimensions of importance** in categorizing harms and risks online
  - Understand the **methodology** of how researchers have developed categorization frameworks
  - Understand differences between **disciplines** e.g., cybersecurity versus public policy

bath.ac.uk
**Descriptive Results**

**Papers per Discipline**

- Misc: 1
- Nuclear: 2
- Business: 5
- Policy: 3
- Social Science: 2
- Information Systems: 10
- Computer Science: 15

**Papers by Year**

- 2005: 2
- 2008: 1
- 2009: 1
- 2011: 1
- 2012: 1
- 2013: 1
- 2014: 1
- 2015: 1
- 2016: 2
- 2017: 2
- 2018: 2
- 2019: 1
- 2020: 2
- 2021: 2

bath.ac.uk
Greater focus on the perpetrator and vulnerabilities of the technology.

Greater focus on the vulnerability of the victim and how they interact with technology.
**Original Aim**

- Aimed to discriminate between risks and the harmful impacts of the risks

- However:
  - Difficulty assessing when ‘actual harm’ has been caused; some harms are anticipatory
  - The link between actions and harm caused are very rarely direct
  - Many ‘nuances in the context’ that may determine the possibility of harm
  - Does there need to be a victim for something to be harmful? The role of awareness in understanding harm
  - Subjective definitions of harm
    - Feinberg’s harm principle states a harm is a ‘setback to one’s interests’ (1984)
    - Harm is a normative judgment that reflects underlying social judgments about good and bad (Lin, 2006)

*bank.ac.uk*
What about risk?

- Taxonomies demonstrate there is no ‘best practice’ way to categorise risk
- Research on risk is reactive and retrospective
- Anything could constitute a risk; surveillance could be risky, adoption of virtual reality technology could be risky, cryptocurrencies could be risky, using social media could be risky…
Vulnerability

- Not all ‘risks’ lead to a harm – and likelihood of harm being experienced varies across people/time/contexts/technologies

- A combination of factors *might* increase the likelihood that a harm is experienced as a consequence of an event...
  - Low trust in institutions makes a country vulnerable to misinformation
  - Shared passwords make a person vulnerable to data loss

- Mitigation works by addressing vulnerabilities at multiple levels/points of interaction
Vulnerability

Two types of vulnerability: outcome and contextual (O'Brien et al, 2007).

‘Outcome vulnerability is considered a linear result of the projected impacts of climate change on a particular exposure unit’

Current focus of much online harms research: ‘risk-hazard’ approach – attempting to understand the impact of a hazard based on exposure to it and sensitivity of the exposed entity (Turner et al., 2003).

‘Contextual vulnerability is based on a processual and multidimensional view of climate–society interactions. [...] From this perspective, reducing vulnerability involves altering the context in which climate change occurs’

What we should be thinking about: modelling the online context and how this system of interacting components leads to contextual vulnerability.
Contextual Vulnerability
Online

bath.ac.uk
The Local Environment

Local Environment Interactions:

A: How do features of the user interact with features of the information in a way that creates vulnerability?
- User supplied information risks: what data is being collected or produced by a user?
- User access to information risks: what content is being accessed or viewed by a user?

B: How do features of the user interact with features of the technology in a way that creates vulnerability?
- Decisional interference risks: how is the technology impacting the user?
- Improper usage risks: how is the user using or exploiting the technology?
- Accessibility risks: Is the user unable to interact with the technology or features of the technology?

C: How do features of the information interact with features of the technology in a way that creates vulnerability?
- Storage risks: how is data being stored by the technology?
- Processing risks: how is data being processed by technology?
The local environment exists by virtue of the commercial environment.

Commercial factors may impact the individual factors of the local environment, or they may impact the interactions between them.

We can break each commercial factor down into its components, for example looking at the impact of each team in a company’s workforce, or each revenue stream in the business model.
The commercial and local environments exist within an overarching societal context.

The societal context impacts the abilities and aims of the commercial environment, which in turn impacts the local environment.

There are also arguments that this relationship is also reversed: the local technological environment influences social and cultural aspects of a society (Postman, 1988).
Interacting commercial or local environments

There may also be non-hierarchical interactions between environments.

For example, the company Facebook (or Meta..) owns Oculus, Instagram, Whatsapp and Facebook. Vulnerabilities are likely to arise from the interaction between these local environments.

There may also be cross platform interactions; how does the local environment of Twitter interact with the local environment of TikTok? How does the interaction between commercial environments lead to vulnerability?
Purpose of this framework

- Identify the interactions between components in the context that may create or cause vulnerabilities
  - This allows us to pinpoint where mitigation strategies could be employed

- Illuminate the possible role of higher order factors, such as culture or legal regulations, in influencing the vulnerability of technological systems

- Anticipate future risks or vulnerabilities in new technologies
  - By defining the features of each environment, we can identify unforeseen risky interactions between system components.

- Move away from presuming direct causal relationships between risk and harm; applying a systems-level approach to the study of online harm.

bath.ac.uk
Thanks for listening.

Happy to take any questions, comments, criticism or compliments 😊

ac974@bath.ac.uk
aj266@bath.ac.uk
ls487@bath.ac.uk
daе30@bath.ac.uk
pssds@bath.ac.uk

@aliciagcork
@joinson
@lauragesmith
@davidaellis
To learn more about REPHRAIN, our future plans and how to get involved:

- [www.rephrain.ac.uk](http://www.rephrain.ac.uk)
- [@REPHRAIN1](https://twitter.com/REPHRAIN1)
- rephrain-centre@bristol.ac.uk

We would love to hear from you. Thank you!

bath.ac.uk