

How Edited and Disappearing Content Poses a Challenge to The UK's Online Safety Regulations Tackling Harm Facing Young People Online

Ellie Colegate | ellie.colegate@nottingham.ac.uk

University of Nottingham, PhD Candidate

School of Law and Horizon Centre for Doctoral Training



Background

- Two timelines are created when content is published online.
 1. Attached to the content and its existence.
 2. Attached to the impact on users exposed to the content.
- If mapped, these could benefit content moderation and regulation when content has the potential to harm users, especially young people.
- Recent changes have refocused the attention of users as to how they can edit content.
- Changed content can limit the ability to document online harms and can cause regulatory challenges.

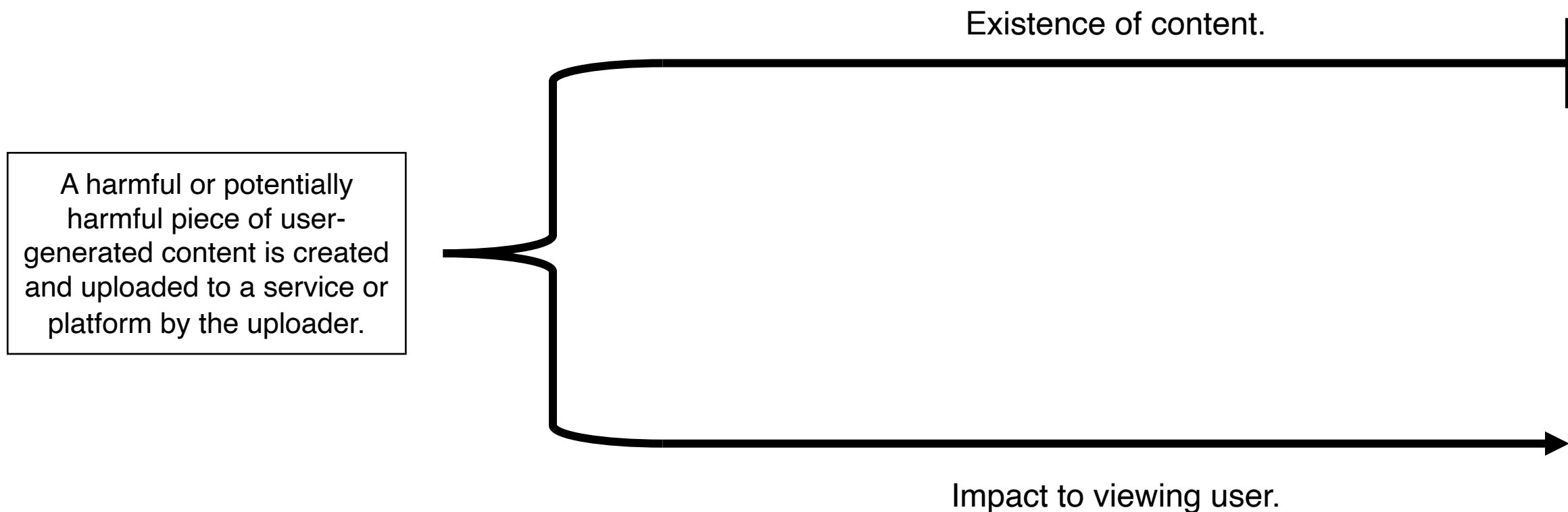


Content Types

- **Section 55 Online Safety Act** - “generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, and that may be encountered by another user, or other users, of the service by means of the service.”
- Within this paper:
 - Retrospectively edited content
 - Content that disappears by design

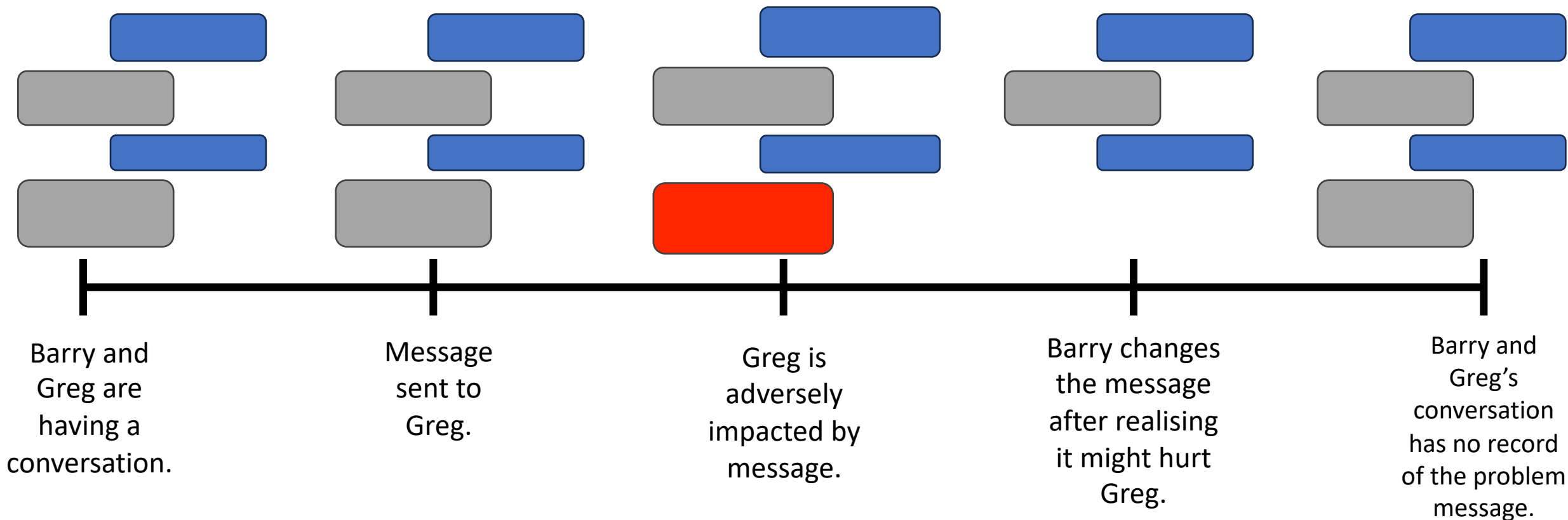


Timelines Related to Content



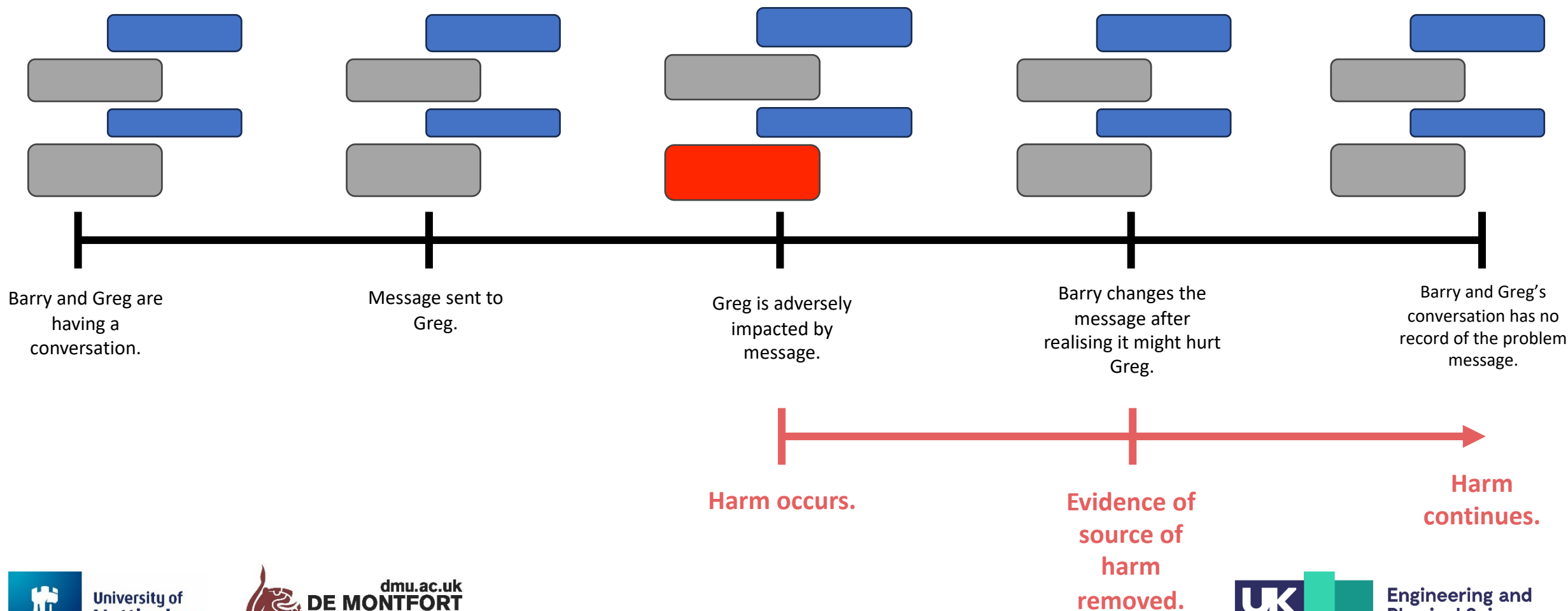


Challenges Presented to Regulation





Challenges Presented to Regulation





Challenges Presented to Regulation

- **Existence of Legacy Content**

- When a user captures the original unedited piece of content and shares it in a new format.
- Real-time transient content can be externally captured and exist beyond its intended lifespan.

- **Possibility of delayed harm to users with no evidential basis**

- Users are not always immediately harmed by what they see online.
- A platform cannot be mandated to remove potentially harmful content that is problematic to users if it no longer exists.
- Unclear accountability for delayed harms stemming from content.



Proactive Platforms?

- **Section 63** – “presents a material risk of significant harm to an appreciable number of children in the United Kingdom.”
- To remove content, a platform must identify the content with the potential to cause harm.
- The ability of users to retrospectively edit or amend content poses a significant risk to a platform's ability to identify.
- Raises questions as to the extent a platform needs to actively adhered to it's online safety obligations.



Ways Forward

Recognition of 'Legacy Content'

- Ofcom acknowledges platform design choices that impact user safety.
- Recognition of the underlying features, such as 'disappear by design' and 'retrospective editing', would capture the impact in relation to online harms.
- A duty to minimise the risks associated with design choices could sit alongside the duties already outlined to reduce the presence of content likely to harm young people.

Mandates to Retain Content Versions

- Periodic reports are a main source of information for trends in content causing harm and harms experienced by young people.
- A move towards transparency reporting by Ofcom, a move towards transparency of evidence?
- A mandate to keep copies of harmful content could provide insight into how features platforms offer can cause harm and are used by users.



Conclusions and Further Works

1. A user's ability to go back and edit content – changing a narrative – could potentially erase the records of harmful interactions.
2. The lack of evidence of harmful interactions can impact stakeholders' ability to identify trends in content that may harm and take pre-emptive action.
3. No content that can cause harm should be online, but we want the internet to retain a record to enable preventative actions by stakeholders.

Thank you for listening.

Ellie Colegate | ellie.colegate@nottingham.ac.uk

University of Nottingham, PhD Candidate

School of Law and Horizon Centre for Doctoral Training

This work was supported by the Engineering and Physical Sciences Research Council Grant No. EP/S023305/1.