TRUST AND ABUSABILITY TOOLKIT:

Centering Safety in Human-Data Interactions
Centering Safety in Human-Data Interactions
Centering Safety in Human-Data Interactions
Centering Safety in Human-Data Interactions
Centering Safety in Human-Data Interactions

authors

Angelika Strohmayer, Julia Slupska, Rosanna Bellini,
Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

This toolkit was funded by UK Research and Innovation's Engineering and Physical Sciences Research Council's Human-Data Interaction Network+ (EP/R045178/1 Human Data Interaction: Legibility, Agency, Negotiability) and parts of this work were funded by a UK Research and Innovation Citizen Science grant (BB/T018593/1).

PUBLISHED IN OCTOBER 2021

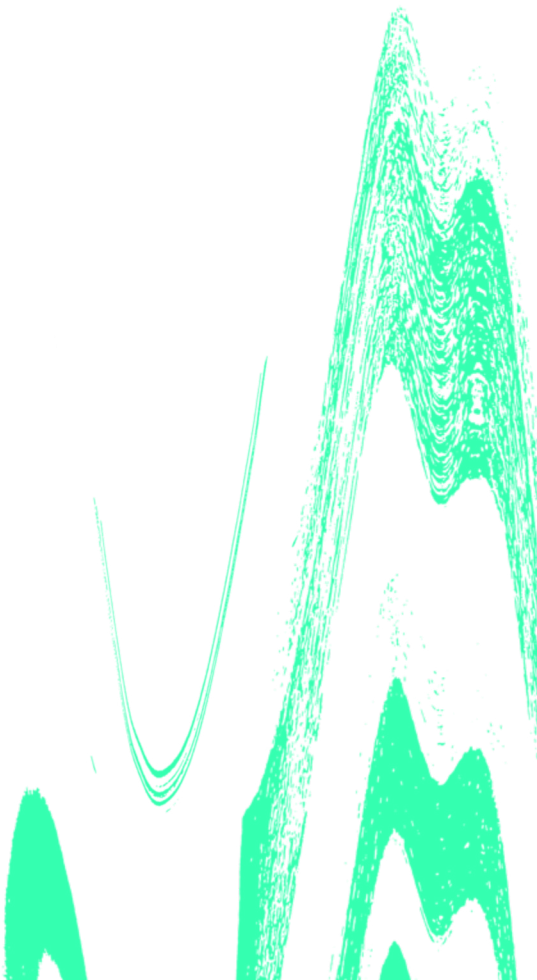
Throughout the toolkit, we have used the Redaction typeface, originally designed by Jeremy Mickel, under Titus Kaphar and Reginald Dwayne Betts for The Redaction exhibition.

This toolkit would not have been possible to make without the expert knowledge shared with us by our research participants and collaborators. We list those who gave their express permission to be named below, in alphabetical order:

- **Adam Dodge**
- **Andrijana Radoicic**
- **Anonymous**
- **Anonymous**
- **Eleanor Goodman**
- **Elissa Redmiles**
- **Emily Tseng**
- **Emma Pickering**
- **Farah Sattar**
- **Honza Cervenka**
- **Jessica (a pseudonym)**
- **Kate Worthington**
- **Leonie Tanczer**
- **Lesley Nuttal**
- **Nadia Khaliq**
- **Nova Ahmed**
- **Rebekah**
- **Sharifa**
- **Tara Hairston**
- **Toby Shulruff**

Design and artwork by Maria Munguambe

Recommended Citation Format: Strohmayer, A., Slupska, J., Bellini, R., Neff, G., Coventry, L., Hairson, A., Dodge, A. (2021) Trust and Abusability Toolkit: Centering Safety in Human-Data Interactions



Executive Summary

If you care about security, you care about safety. So you need to care about abusability and trust. This toolkit will provide information about why centering peoples' safety in our digital technologies is important. We present the concepts of abusability and trust as two important tenets of building such safer technologies, followed by resources that can help us build safer technologies.

How and why we trust in technologies can have many different meanings depending on the context. For example, we can understand trust to be a static object (eg. when a trusted user is given access to a system), but we can also understand trusting as an action (which is impermanent and can waver in time) and being trustworthy as a quality to aspire to. Despite this complexity, when it comes to designing new technologies, we often only consider trust as a shorthand for user engagement - if we increase trust in a system, more people will use it. We argue that when we center peoples' safety instead of engagement, trust has to become a more complicated part of the development process. Instead we provie three questions that can help designers and developers more deeply engage with the topic of 'trust' and how it can help us build safer systems: (1) Is trust being used as a verb or a noun? (2) Who, where, and why do we Trust? And (3) how do we Trust?

Abusability is the possibility that malicious actors might weaponise a system for harmful activity; designers have a responsibility to anticipate and mitigate this. Abusability plays on the concept of "usability" to ask what kinds of uses should be restricted rather than enabled in design. This form of threat assessment and testing reframes security's traditional focus on an external attacker penetrating or subverting a system from its intended or authorised purpose, to ask how people use features for harm. Developers and designers should consider abusability at various stages of the design lifecycle: asking how might a product be abused during threat modelling and testing, and then making sure there are robust mechanisms for responding to abuse, such as reporting features and support for rectification. This allows companies to respond to problems on their devices and platforms as they arise. Mature company practices in this space will move beyond tokenistic inclusion and meaningfully involve survivors and advocates in designing both preventative measures and responses to abuse.

A summary of the resources available at the end of the toolkit

- **Implications for advocates and others who support survivors** outlines key recommendations for supporting survivors of technology-mediated abuse and for engaging with technology companies to improve their services
- **Designing for Survivors and Perpetrators** outlines advice for technology companies and researchers wanting to implement features that better support survivors of technology-mediated abuse and proactively engaging with perpetrators of such harm
- **Abusability and the secure systems development life cycle** provides an outline for a development life cycle that takes peoples' safety into consideration
  We provide three **Case Studies** that can be used to explore, learn, facilitate training, or help people engage with the topic of safety when designing new technologies
- **Taking Abusability Seriously** provides technology companies and services a self-evaluation tool of how mature their features are in relation to the safety
- **Do's and don'ts for journalists covering technology-mediated abuse** provides guidance and advice for journalists who write about topics related to technology-mediated abuse

Key take-aways for technology workers and companies

- You cannot 'design out' harm and abuse, but you can reduce the likelihood it will happen and mitigate its effects through responsive company processes
- Abusability is something you have to consider throughout the design and deployment processes, and should be negotiated with people who use the system after it has been deployed
- Trust is dynamic, and technologies can encourage us to examine new questions about ourselves and our connections with others; and we should have the confidence to ask questions of such devices.
- Focusing more on the process of designing new systems or features (rather than being entirely outcome-driven) can help us be more careful about what we design and the potential negative consequences it can have

Key take-aways for advocates, support workers, and survivors

- Your expertise is invaluable and should be included in the entire design cycle of a project
- Be wary of 'safety washing' where companies involve you only at the end of a project to get 'approval' and be able to write that they worked with advocates / survivors

Key take-aways for academics, researchers, and facilitators of learning

- When thinking about human-data interaction, it is important to consider how power plays into the conversation, what the wider eco-system in which the interaction sits includes, and to be more process- rather than outcome-oriented in our work
- This toolkit is a resource for students and facilitators of learning for computer scientists, designers, software developers, machine learning training, etc. Please make use of it as such.
- Feminist ways of working can influence not only waht we do, but also how we approach the topic/issue/problem. It is important to change what we see as 'security' and 'safety' concerns in technology research, and how we are able to deal with these issues holistically rather than attempting to design individual systems or features that aim to end this complex, societal problem.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

About the authors

Dr Angelika Strohmayer is co-director of the Design Feminisms Research Group and senior lecturer at Northumbria University's School of Design. She is interested in developing justice-oriented and feminist theories for technology research, one strand of which focuses on issues of safety and harms.

Julia Slupska is a doctoral student at the Centre for Doctoral Training in Cybersecurity and the Oxford Internet Institute. Her research focuses on technologically mediated abuse like image-based sexual abuse ('revenge porn') and stalking, as well as emotion, care and metaphors in cybersecurity.

Dr Rosanna Bellini is a postdoctoral researcher for the Technology and Intimate Partner Violence research group at Cornell University, formerly at Open Lab at Newcastle University. She specialises on the role of technology in interventions for perpetrators of domestic abuse, and in documenting attack vectors for technology-facilitated abuse and intimate partner surveillance.

Prof Gina Neff is Professor of Technology & Society at the University of Oxford and the Executive Director of the Minderoo Centre for Technology & Democracy at the University of Cambridge. Her books include Venture Labor (MIT Press 2012), Self-Tracking (MIT Press 2016) and Human-Centered Data Science (MIT Press 2022).

Prof Lynne Coventry is a Professor in Psychology. She leads the Human & Digital Design multidisciplinary research theme across Northumbria University. She has an interdisciplinary human computer interaction background. Her research focuses on inclusive interaction with technology ensuring diversity, of needs and abilities, are accommodated through the design.

Tara Hairston is the Executive Director of the Coalition Against Stalkerware, a volunteer group of over 40 organizations globally, dedicated to raising awareness of stalkerware as an example of technology-facilitated abuse, providing resources to assist survivors and other individuals targeted by it, and working with stakeholder groups to better detect and mitigate against these surreptitious surveillance tools.

Adam Dodge is the founder of EndTAB (End Technology-Enabled Abuse). His work focuses on training victim-serving organizations to prevent and address the ways in which people are harmed online and via their devices.

Definitions

Human-Data Interaction (HDI)

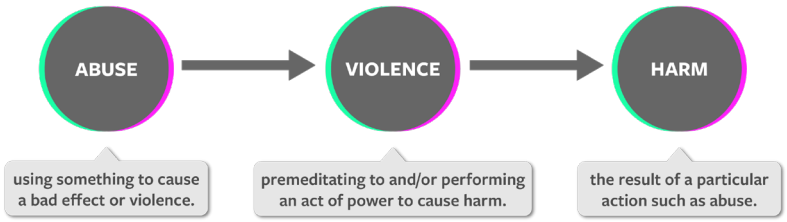
Human-Data Interaction (HDI) is a framework that presents three tenets that help designers make more ethical data-intensive systems. Data intensive systems are digital infrastructures and systems that use, curate, or analyse data that is produced by and about us as people. HDI tries to make sure that this data (1) is legible (which means we have access to it in a format that we are able to understand). (2) HDI also postulates that we should have agency over how data is collected and used about us; and (3) that we should be able to change how this data is used by negotiating with those who receive data about us.

Data-Intensive System

Throughout the toolkit, we refer to devices that collect, produce, or analyse these kinds of data that we produce as 'data-intensive systems'. These may be algorithms or smart home devices that rely on algorithms and machine learning to do their computations - for example a smart speaker or a smart watch. Data-intensive systems however not only refer to objects or 'things' that we can see, but also relate to software systems. For example, in the UK a data-intensive system was used to calculate A-level results in 2020 to great uproar from students, teachers, and parents alike.

Violence, Harm and Abuse

By **violence** we mean the intentional use of power, threatened or actualised, against yourself, another person, or a group that results in injury, harm, maldevelopment or deprivation. This is similar, but not identical to a more commonly used term of **abuse** that can be defined as using something for a bad effect or purpose and treating someone with cruelty or violence. As many people only associate 'violence' with physical harm, many prefer the term 'abuse' though you may also see this language being used interchangeably. The final term that this toolkit covers is **harm**, which is the damage caused to someone caused by a particular course of action. As we can see each of these terms builds on or directly includes the other, and we can roughly map them out as so:



Trust

Trust has many different meanings in different contexts, but for simplicity we use the definition proposed by Marsh and Dibben (2003): trust includes a positive expectation (that something will happen) regarding someone or something's behaviour in a situation that involves risk. This can mean looking beyond what feels trustworthy subjectively and onto analysing larger group behaviours such as what it means to trust in software companies in resolving challenges due to malicious or absent-minded intentions. Trust is especially important in cases of abuse and violence as abuse is not only a misuse of the positive benefits accrued from being in a position of trust, but is an integral part of trust itself.

Technology-mediated abuse

This can sometimes be referred to as technology-facilitated abuse or 'tech abuse', but includes any deliberate use of digital technologies or systems to scare, harass, coerce or stalk someone. Examples include intimate image abuse (sometimes known as 'revenge porn'), using location-data for stalking, or remotely controlling Internet of Things (IoT) devices for 'gaslighting', i.e. making someone doubt their own sanity. It is a form of controlling behaviour as its intention is to control or influence the behaviours of the person(s) who are being targeted. It can result in someone feeling helpless, isolated, and confused due to the pervasive and ubiquitous nature that technology plays in modern life. Like all forms of abuse, technology-enabled abuse disproportionately affects those whose identities are marginalised due to societal structures of oppression such as racism, misogyny, class privilege, ableism, heterosexism.

Introduction

As people who use smartphones, computers, or other smart devices, we interact with and produce data on a daily basis. This data is then used by companies, work places, or organisations for a variety of different goals. In this toolkit we present pragmatic opportunities for developers, designers, and third sector organisations on how we can interact with these kinds of data-intensive systems in ways that center peoples' safety. We do this by first presenting the Human-Data Interaction Framework as a useful set of guiding principles for ethical engagement with humans and data, before we extend the existing framework with pragmatic approaches to how we can more directly integrate notions of 'abusability' and 'trust' into the design, deployment, and use of such systems. At the end, we also present a number of resources to help implement our ideas.

By data we mean a record of something that has happened (a heartbeat, a mouseclick) that has to undergo a transformation through analysis or observation into information. The buzzwords 'data-driven' and 'data-informed' policy and approaches have become commonplace but there has been little confirmation as to what these actually mean for different groups and contexts. In doing so, this brings about major opportunities for the design of different processes and systems that have the potential to enhance our lives. Information can be used to learn new insights about people, places, and important concepts. Whenever we get a change of state, from data to information, we can ask the following questions: What is chosen to be useful to collect information on? Who is included or excluded in this transformation of data to information? What aims do we hope to reach through having such information? What information is lost in the process of analysis? Asking these questions and responding to them has the ability to expose injustices, particularly for individuals who do not have the ability to ask these of data-intensive systems and the people or companies who make them.

All of these questions mirror some interesting conversations around the distribution of social power across different social groups and identities. Social movements such as #MeToo and BlackLivesMatter make evident, among other things, the power imbalance in crucial social systems upon which we depend, such as employment and right to justice. Feminist thinking and theories give us language to critically analyse these power imbalances, particularly looking towards how women and other marginalised groups may not hold power over important decisions that affect their lives (Hendricks & Oliver, 1999). Because of this, it is essential to scrutinise frameworks that help us develop more ethical structures for how humans interact with and through data and data-intensive systems. With the move towards decision-making systems based on so-called Big Data, there has never been a better time for a call for critical action and analysis.

There is often a huge gap between developers of such systems who want to work ethicallyu, and the tools needed to be able to put this willingness for ethical conduct into practice. While there are many frameworks for 'ethical design' or 'ethical data' practices, many of these are very conceptual. This means they do not necessarily provide computer and data workers with pragmatic guidance on how to put the well-meaning conceptual ethics frameworks into practice.

The Human-Data Interaction (HDI) Framework was developed in 2016 by a group of researchers working across a variety of fields to fill this gap and to help computer and data workers engage pragmatically with ethical design practices around data-intensive systems and the services they provide. When thinking about 'systems' we are referring mostly to computer systems (both the hardware and software that are needed for it to function), but

depending on our projects, this may also relate to wider systems of who is using these computers, in what settings, and for what purposes. So for example, in an office you may be using a laptop with a variety of software systems installed. These systems may be related to different forms of data that your workplace collects, analyses, and uses in spreadsheets or text documents. In this case, it is important also to think about the setting you are in when using the laptop computer - your access to certain kinds of data probably depend on what department you are in in your company, and where you are on the corporate hierarchy.

These different kinds of 'context' are important to consider, because our definitions of 'system' really depend on the scope and frame of our analysis. This is also the case when we think about the HDI framework, though for the most part, we refer to data-intensive software systems that predominantly deal with personal data. This relates for example to algorithms or smart devices, and how they compute and use data that is collected about us as individual people.

When designing such data-intensive systems, designers and developers always have to make decisions about what kinds of data to collect and analyse, how to represent them, and how to reduce and abstract them for the systems to work. It is not a question about whether reductionism or abstraction take place in systems, but rather about how and why that takes place. Computer systems are media like many others, where designers are able to emphasise, aggregate, or ignore different kinds of information; they make choices about how and what to abstract or reduce - it is not possible to not do this. As researchers who have come before us have said (Chalmers, 2004a, 2004b), it is not possible to take everything into consideration all of the time.

The HDI framework provides us some starting points to think about this issue in more detail. It provides us some steps towards building better systems that are more ethical in their use and computation of data. We hope that throughout this toolkit you will learn about the HDI framework and how it can be helpful - we also hope that our critique of this framework and its extension will help you be able to better design data-intensive systems that take into consideration peoples safety. For clarity, we, as authors of this toolkit, did not create the HDI framework - we were funded by the EPSRC Network+ on Human-Data Interaction (which is a group of academics, some of whom were part of the team who coined the term and developed the framework) to explore, critique, and extend the existing HDI framework to center peoples' safety.

RESEARCH METHODS

The work presented in the toolkit took place in the Spring and Summer of 2021, using a variety of methods in a feminist framing of safety and digitally-mediated abuse of data-intensive systems. Here, we first outline the methods we used for our work individually before we address their convergence in the development of this toolkit.

Research workshops with experts:

Julia Slupska and Angelika Strohmayer, with support from Tara Hairston, Gina Neff, and Adam Dodge, worked to critique and extend the existing Human-Data Interaction (HDI) framework. We facilitated two workshops:

1. A workshop with 11 international expert practitioners who work on topics related to safety, data, and technologies. This included third sector organisations who support victim survivors, security and technology practitioners, as well as lawyers and human rights advocates.
2. A workshop with 9 international expert researchers who work on topics related to technology-mediated abuse, security, other forms of violence related to technologies, and data justice. This included academics at various career stages as well as researchers from industry.

X

Our feminist lense relates to our critical engagement with the data and the HDI framework overall. This relates specifically to how we center experiences that are usually seen as 'edge cases' such as experiences of intimate partner violence and technology-mediated abuse. It also refers to our underlying aim of centering peoples' safety in the development of new, and bettering of existing, data-intensive technologies. As part of this, we reflect on power structures, explore assumptions from the margins, and unpick concepts beyond their immediate meaning. Through our research methods, we strive to include multiple and disparate voices, bringing together different types of expertise in an equitable manner.

These workshops were audio recorded and transcribed. The transcripts were then analysed using Reflexive Thematic Analysis (Braun and Clarke, 2020) through a feminist lense - an approach that provides a reflection of qualitative research as "creative, reflexive and subjective, with researcher subjectivity understood as a resource (see Gough and Madill 2012), rather than a potential threat to knowledge production." (Braun and Clarke, 2019). We followed Braun and Clarke's (2006) 6 phases of doing a Thematic Analysis: (1) data familiarization, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, (6) and producing the report.

Literature Review on Abusability:

At the same time as the thematic analysis, and after initial coding of the workshop transcripts and multiple group reflections on initial findings, Julia Slupska carried out a literature review to collate articles that relate to feminist methods for the development of databases and data-intensive systems, focusing on project processes and narratives of power, and the concept of abusability. Ultimately this analysis brings together both empirical and theoretical research to explore the relationship between these disparate areas of study. She did this to find out how design processes, power structures, and abusability relate to the development of computer systems. This process was also framed by Slupska's previous research with advocates of victim survivors, as well as the literature that was shared during the workshop with expert researchers on related topics. Strohmayer wrote up Slupska's initial analysis.

Narrative Literature Review on trust, violence, and technologies:

Rosanna Bellini, under supervision from Angelika Strohmayer and Lynne Coventry, carried out a rapid evidence review by following and adapting the 26 recommendations for such processes by the Cochrane Research Group (Garritty et al, 2021). This review explored how 'abuse' and 'trust' are depicted and framed by Human-Computer Interaction (HCI) and other related technology literatures. This approach to the literature review allowed her to do a comprehensive review of the literature, while also giving her space to explore broad and interdisciplinary articles. The analysis was done not only to critique current discussions, but also to reflect on how the work of the academic community is embedded in wider conversations around violence. Rosanna selected, parsed, and analysed 110 articles for the review, developing a dynamic and relational framework to help us better understand what 'trust' means in relation to technologies and abuse.

Producing Case Studies:

Following a series of interviews with advocates who support survivors of technology-mediated abuse, a workshop with additional advocates, and the production of a literature review on abusability, Julia Slupska developed a series of case studies to illustrate what we mean with 'abusability' and 'trust' in our extension of the HDI framework. This was carried out in conversation with Angelika Strohmayer, and brings together examples from the empirical research with advocates as well as examples from industry. The case studies were produced to illustrate the complexity of the real world, and how abusability and trust can help us develop tactics and technologies to improve safety for people interacting with data-intensive systems.

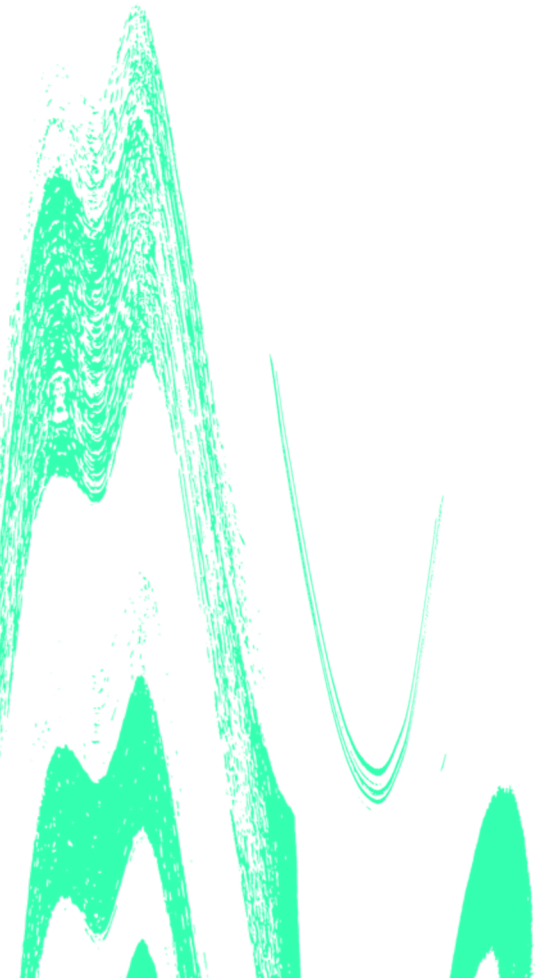
Creating the toolkit :

This toolkit brings together all the research processes described above. We of course also draw on past research and work experience from the authors, giving us space to situate and contextualise our analyses (Braun and Clarke, 2020). As such, the resources shared at the end of the toolkit were created based on the research methods we outline above, but also stem from knowledge produced in a variety of other research projects carried out by the authors over the last years.

OVERVIEW OF THE TOOLKIT

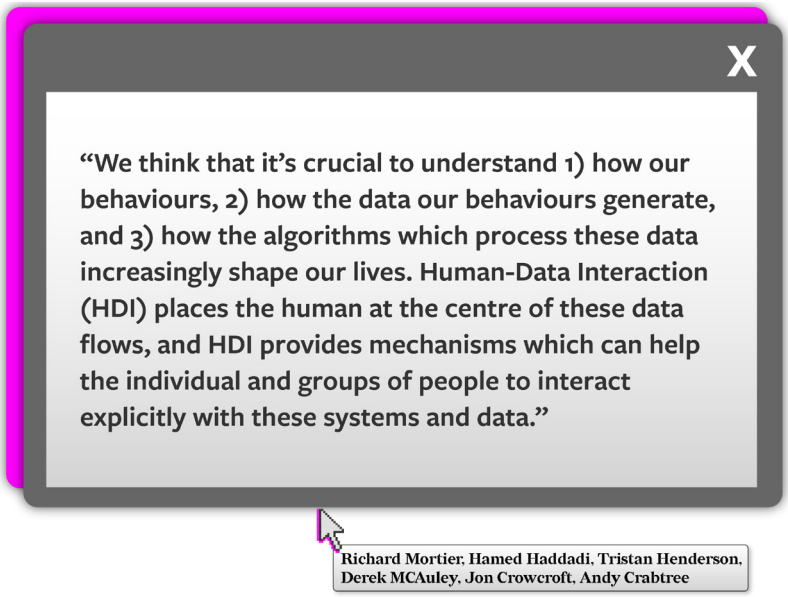
The toolkit is laid out in three parts: (1) an introduction to the Human-Data Interaction (HDI) framework; (2) a feminist critique and extension of the existing HDI framework that provides insight into how we can better center peoples' safety in data-intensive systems; and (3) a collection of resources that can be used by technology designers, advocates for victim-survivors, educators, journalists, and others in this area of work to develop better technologies and systems that place peoples' safety at the heart of their work.

Throughout the toolkit, we present critiques of current approaches related to the 'ethical' use of data while also producing new and powerful insights that can help us build better systems and technologies in the future. We do this by focusing our critique on processes rather than outcomes, giving space to critically examine and explore alternative ways in which we can engage people and threats in product design cycles. We also look at how these technical systems and their design processes can help shape societal change (looking towards social education, funding landscapes, and policy discourse) to tackle the root causes of digitally-mediated abuse and violence.



Part 1:

WHAT IS THE HUMAN-DATA INTERACTION FRAMEWORK?

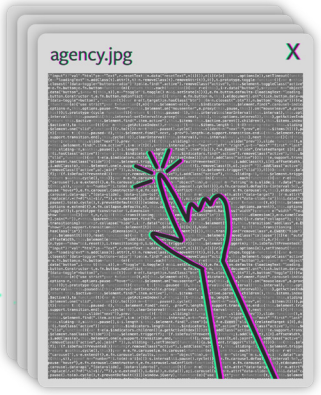


In 2016, Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek MCAuley, Jon Crowcroft and Andy Crabtree wrote about Human-Data Interaction for the Encyclopedia of Human-Computer Interaction. In this article, they wrote: "We think that it's crucial to understand 1) how our behaviours, 2) how the data our behaviours generate, and 3) how the algorithms which process these data increasingly shape our lives. Human-Data Interaction (HDI) places the human at the centre of these data flows, and HDI provides mechanisms which can help the individual and groups of people to interact explicitly with these systems and data."

However, this is not the first or only definition of this framework. For example, publications from 2013, 2014, and 2020 produce a number of slightly different definitions. However, we use the definition presented on the HDI NEtwork+ website (https://hdi-network.org/) that outlines pragmatic guidance for the data-intensive systems development process: legibility, agency, and negotiability.

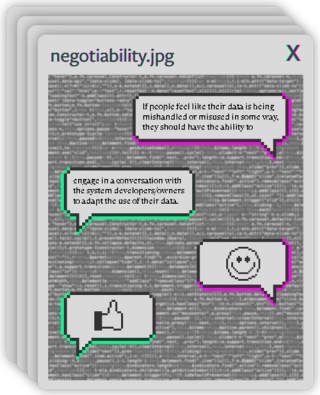
Legibility

Legibility refers to the understanding that data is not only presented in a transparent way, but that the way it is presented is understandable to people. This means it is not only possible to look up what data is being collected, analysed, or used about people in a certain system, but that people are also able to understand what it is they are looking at. This requires a certain degree of curation or simplification from the system developers/owners.



Agency

Agency refers to the ability for people to have the capacity to interact with the system in which data about them is being collected, produced, or used. In short, people should have agency over what data is collected and stored about them, how it is used, and what happens with it.



Negotiability

Negotiability refers to the ability for people to not only have agency over what happens with their data, but that they are able to negotiate this use with those who run the system they are using. If people feel like their data is being mishandled or misused in some way, they should have the ability to engage in a conversation with the system developers/owners to adapt the use of their data.

To build on this thinking, to develop more pragmatic principles, and to share examples of how this way of thinking could look in practice, Matthew Chalmers and his colleagues developed the Human-Data Interaction Network+ in 2018. Chalmers and his colleagues understood that there was a conceptual gap between software development work and wanting to do the 'right thing' pragmatically to ensure peoples' data is used ethically. They appreciated that it is important to think about not only the direct software development, but also to think about the contexts in which and with which these developers work. For example, what code and data libraries are being used, who are the people developing these data-intensive systems, or what are the corporate and time pressures under which people have to work.

As part of the Network's work, they extended the framework, adding the concepts of '**Surveillance/Resistance**'. While not formally added yet, this recent addition provides further thinking of how people can interact with data and data-intensive systems - a way of resisting use of data that is perhaps legal but not seen as 'ethical' by the people about whom data is being collected. This is a concept that should be used when the first three concepts (legibility, agency, and negotiability) are already applied, but we still believe our data is being mishandled (even if legally so). While there are many examples of 'defensive' approaches to resisting the use of personal data by corporations or governments, there are few examples of techniques that make the perceived misuse of data visible through the use of the system (eg. by increasing noise in datasets or adding false data as activism). These are the kinds of 'resistance' that Chalmers and his colleagues refer to.

Essentially, when taking these four concepts together, we should have certain kinds of control over data-intensive systems that collect, use, analyse, or share our data. We should be able to understand what is collected about us and how this is used (legibility), we should be able to make decisions about what data is collected and how it is used (agency), we should be able to raise concerns about perceived misuse of data with the ability to make changes to the system (negotiability), and we should have ways of resisting to legal but morally dubious ways of using our data (surveillance/resistance).

WHY IS THE HDI FRAMEWORK NOT ENOUGH TO ENSURE PEOPLE ARE SAFE?

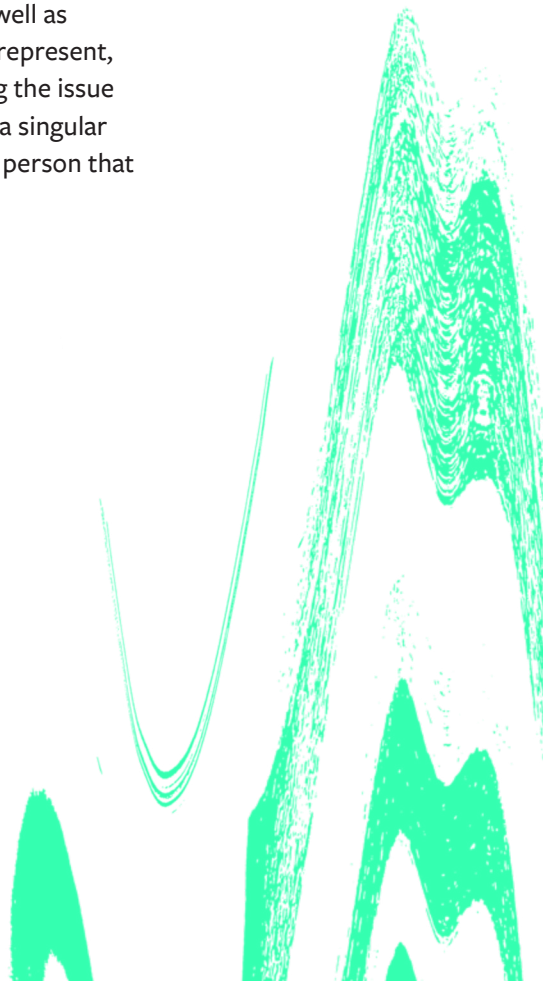
Despite its potential, the HDI framework is not enough for us to be able to pragmatically provide ethical and safe technologies. We go into details of this in Part 2 of this toolkit, but want to present a brief overview here as well. Taking a feminist lense to HDI, we can quickly see that the framework is not complete. It lacks a deep engagement with power structures, how our individual personal data relates to the personal data of others, and how all of these concepts have implications for people's safety, especially those that exist at the margins.

To illustrate what we mean, we now take a closer look at the framework's concept of 'agency'. While it is important to give people the capacity to interact with systems and to control and correct data-driven processes, not all uses of agency over data are beneficial. Malicious uses of this agency can contribute to controlling behaviours in interpersonal relationships, for example as is shown by Leitão's (2019) work anticipating the abuse of IoT devices with domestic violence survivors. There is also not enough critical discussion in the HDI framework, and elsewhere, of when and how agency could and should be limited or constrained.

To carry out our feminist critique, we center experiences and needs of those people who are experiencing different forms of oppressions from systems, society, or legal frameworks and as such have unique ways of interacting with data and data-driven systems. In traditional security research, these are often referred to as 'edge cases'. These 'edge cases' however make up huge swathes of the population. To name only a few examples, social media algorithms shadowban activists which can result in harming movements towards more socially just worlds (Blunt et al., 2020), and we know that data-intensive systems have deeply ingrained gendered and racialised biases (Eubanks, 2018; Noble, 2018; Costanza-Chock, 2019). Data-intensive technologies are also being used as part of other forms of violence and coercive control such as stalking, and in domestic violence aggressions (Freed et al., 2017, 2018, 2019; Woodlock, 2017; Tseng et al, 2020; Bellini et al, 2021). Experiences of digitally mediated violence, such as algorithmic bias or shadow banning, become even more commonplace when people experience multiple, intersecting, and complex forms of oppression (Blunt et al., 2020; Crenshaw, 1989).

A UNIDIR report (2021) on gender-based cybersecurity explored issues of design, defence, and response to violence perpetrated through digital means has noted that criminal justice responses are currently not adequate in protecting victim-survivors and holding perpetrators to account. Normative language in security discourse often relates to victim blaming (eg. “they didn’t do enough to protect themselves”) and results in the individualisation of the problem and responsibility (eg. “these are individual incidences of violence”). Feminist approaches to security emphasize that social relations are a source for both security and insecurity, or a “key connective tissue through which different dimensions of (in)security are entangled” (Hörschelmann and Reich 2017).

Furthermore, feminist research from criminology and sociology about experiences of domestic violence highlights that violence is always experienced as part of an ecology which is mediated through different forms of violence (verbal, physical, etc.) as well as coercion and control. In recent years, this has also developed into a conversation about digitally-mediated violence (Leitão, 2019; Tseng et al, 2020; Bellini et al., 2021). Understanding our world as post-digital (Coles-Kemp et al, 2020) means that we know that people’s safety is impacted when interacting with data, databases, and data-intensive systems; and that this can impact aspects of both our digital and non-digital selves. We also know that any violence we experience based on this has real, material impacts on our physical, mental, and emotional health. Additionally, seeing safety as central to our post-digital worlds, and seeing this safety as part of the ecologies in which we interact with data, means that we see the impacts of data-driven and data-mediated violence on individuals as well as people around them: their children (Millar et al., 2021) or the communities they represent, through increased racism (Jankowicz et al., 2021), for example. Ultimately, seeing the issue as an ecology of violence allows us to understand that these incidents are never a singular anomaly of use of the data or data-intensive system, and that it is never just one person that this kind of abuse is being perpetrated against.



USING OR RE-DESIGNING TECHNOLOGIES TO END INTERPERSONAL VIOLENCE

As part of this toolkit, we present two potential extensions to the HDI framework - concepts that will help developers and owners of data-intensive systems think more critically about peoples' safety. Our intention is that through this, technologists will be able to develop better (and safer) technologies for all, but especially for those who exist at the margins.

Having said this, we want to very strongly refute the thought that if only we are able to design the right kind of technology or the right kind of feature in our data-intensive systems that we are able to make people entirely safe. Rather, we would argue that this technology-centred approach to the abuse of data and data-intensive systems to perpetrate violence against individuals is in itself harmful. What if we imagined a world where money was spent on support services for survivors of violence or behaviour change programmes for perpetrators instead of unused digital safety gadgets? **Imagine the world we could create if we tackled the issue of abuse and misuse of data-intensive systems at their root; if instead of tackling mis-use of one specific system, we instead tackled issues of patriarchy and of violence at all levels of our societies?**

While we do not see technologies as solutions to abuse, violence, and other societal problems, technologies are indeed useful tools in the wider systems in which we exist. Technologies mediate so many aspects of our lives, that it seems almost inevitable that they also mediate abuse as well - this is an aspect that should be accounted for in design processes, rather than being ignored. Because of this, **it is not possible to 'design out' the abuse of technologies, or the use of technologies to harm others.** However, it is possibly to build systems which are safer and less prone to abuse. We present some reflections on existing data-intensive systems as well as what we can learn from their (mis) use below.

Part 2:

FEMINIST CRITIQUE AND EXTENSION OF THE HUMAN-DATA INTERACTION FRAMEWORK

Part 2 of this toolkit relates directly to the research carried out about the Human-Data Interaction framework. First, we present a short overview of academic literature that sits at the intersection of technologies and harm. Then, Strohmayer writes about how researchers and practitioners have critiqued the framework through a feminist lense. After this extended critique, Slupska and Bellini expand the framework, adding two important concepts that developers of data-intensive systems should take into consideration: abusability and trust, respectively. Following these additions, Slupska presents a series of case studies that represent both positive and negative examples of when developers and users of data-intensive systems have engaged with notions of abusability and trust.

AN OVERVIEW OF THE ACADEMIC LITERATURE

Many researchers have studied the relationships between technologies and harms. Here, we want to give a very brief overview of this work, following a review of the literature. We split this overview into two subsections: (1) empirical research about technologies and abuse; and (2) theories that are used to explore this research area.

Empirical research to study technology-mediated abuse

We found that there was little work about technology-mediated abuse prior to 2018, but that this area of research has expanded very quickly afterwards. During this time, there seem to be two 'waves' of empirical research at the intersection of abuse and technologies: wave one relates to empirical work describing the problem and wave two refers to the development of solutions to support survivors. While we do talk about somewhat chronological 'waves' of research, this is not a mutually exclusive distinction between empirical studies and prescriptive design work, as many projects have elements of both. This review is not all-encompassing and was carried out with the function of developing a structured narrative to help us understand the current landscape.

In the first wave, researchers focused on researching the kinds of attacks that perpetrators of technology-mediated abuse use (Levy, 2015; Henry and Flynn, 2019; Freed et al, 2018; Lopez-Neira et al, 2019, Tseng et al, 2020), survivors' digital privacy and security needs and practices (Matthews et al, 2017; Freed et al, 2018; Dragiewicz et al, 2019; Sambasivan et al 2019, Harris and Woodlock, 2019; Suguira and Smith, 2020) and the role of platforms in mediating abuse (Dragiewicz at al, 2018). Looking towards a different way of doing research, some researchers have also attempted to numerically quantify frequency of abuse experienced across various platforms, both in individual countries (for example, Pew Studies in the US) and across the world (Thomas et al 2021).

Of course this academic research also makes use of some incredibly important grey literature. This includes, for example, reports from women's support services (see eg. Project Shift, 2017; Glitch, 2020; WESNET, 2020), survivor support organisations (Suzy Lamplugh Trust, 2021), sex worker organisers (Barwulor et al, 2021; Blunt et al, 2021 ; PLAN International, 2020). This work often provides more timely empirical research about experiences of harm and violence, as well as opportunities for the development of services to support perpetrators in changing their behaviours. Government reports, such as the Australian eSafety Commissioner's (2020) work on how technology abuse victimises children of domestic violence survivors, have also been critical. Privacy activists like Eva Galperin and the Coalition Against Stalkerware have also made significant contributions to our understanding of combatting stalkerware in practice (Galperin 2019). This work often includes powerful testimonies of survivors, as well as practical and impactful advice on how to make changes in services, policy, or legal structures.

Following this intensive empirical research to understand practices of perpetrators and needs of survivors, researchers in the second wave turned their attention towards finding solutions to this problem (Kadri 2020). This includes, for example, Kadri's (2020) call for greater empathy from technologists, police officers, educators and employers as well as the development of Technology Abuse Clinics (Havron et al, 2019; Freed et al, 2019). And Zou et al (2021) investigate the role of computer security customer support in aiding survivors. This empirical work also relates to the development of design recommendations (Levy and Schneier 2021 IBM 2020, Chayn) and methods such as threat modelling (Slupska and Tanczer 2021), co-design with survivors (Leitao 2019), and usability analysis (Parkin et al 2019) for the design and development of safer systems and technologies as well as legal and policy recommendations (see eg. Citron 2019). There are also examples that have put into practice some of these recommendations, such as Arief et al's (2014) platform for survivors, the Tech vs Abuse (https://www.techvsabuse.info/) project, or Unmochon (Sultana et al 2021).

There are few exceptions of empirical work that studies the relationship between technologies and harms that relates to perpetrators rather than victim-survivors. While it is incredibly important to continue to study ways in which we can better support survivors, it is also important to study the practices of perpetrators (Tseng et al, 2020, Bellini et al, 2021) - and importantly also how their behaviours can be influenced (Bellini et al 2021).

All together, this empirical work of the last eight years has provided detailed information and accounts about the nature of technology-mediated abuse and has provided some insights about how to counter these harms with technologies.

Theories of studying technology-mediated abuse

Researchers have also written more theoretically about the intersection of technologies and abuse. These relate to a diversity of diciplines, such as feminist security studies (see eg Sjoberg 2018) and feminist technology studies (see eg Wajcman 2007).

The relationship of technologies and safety also relates to, for example, the Glitch Femninism manifesto that addresses the ways non-conforming bodies are glitches in technical systems; and how our digital lives and our 'Away From Keyboard' (AFK) lives are deeply intertwined (Russell, 2020). This idea is similar to Coles-Kemp's idea of living in a post-digital world, and that this has specific implications for our safety (Coles Kemp 2020).

These theories give us language and understanding to look at harm and violence perpetrated and experienced through technologies as part of wider ecologies of harm. This means that we understand it is not possible to entirely design-out harm, as we have already mentioned above. But it also means that we have to cast our web wider than the design stage of the development of new technologies when thinking about technology-mediated abuse and its implications for people. For example, feminist theories make it clear to us that not only are outcomes of design processes important, but that the processes in the development and use of these systems themselves can be powerful agents of change.

Furthermore, these approaches give us space to think more deeply about what safety means, and who's safety it is we want to center. All our AFK and virtual experiences are shaped by who we are, the different identities we inhabit and present with, and how others perceive our identities. For example, we know that black women are more likely to experience abuse online than white women, and disabled people experience different kinds of abuse to non-disabled folks. While we of course have to design our systems in ways that reduces the opportunities of harm, we also have to design ways in which people can report or counter harmful situations when they arise - and importantly - that these complaints will be taken seriously and acted upon.

All of this relates to power. The power we have as individual users and in collectives, but also the power that Big Tech has over us - the power of data. This is what brings us back to the Human-Data Interaction framework. In this toolkit, we focus on the issues of human-data interaction, including abuse and harm.

A FEMINIST CRITIQUE OF THE HUMAN-DATA INTERACTION FRAMEWORK

One part of our workshops with academics and advocates who support victim-survivors of technology-mediated abuse constituted critiquing the existing HDI framework. The conversations across the two workshops can be summarised into three key areas: (1) HDI has a lack of power discourse within the framework; (2) There is a need for us to see and understand wider ecosystems in which HDI and support sit, rather than focusing only on the immediate issue; (3) There is a need to be less outcome-oriented and instead to see the process of designing and developing new systems and/or services as equally, if not more important

(1) HDI has a lack of power discourse within the framework.

Our world is underlined by capitalism and patriarchy; and this includes the development and use of digital technologies and data-intensive systems. In its current state, the HDI framework is made up of three (or four, if counting the most recent extension) separate tenets that all design of data-intensive systems should follow. However, these tenets are all based on an individual's interaction with their data through the system - they do not take into account the wider systems in which people and the technologies sit. For example, in today's world, one person's data impacts on others as well - algorithms are built on data that has been collected from many people, not just individuals.

Thinking about HDI as part of a wider system gives us new opportunities of how we can engage with safety concerns and/or the abuse of systems. For example, one participant said: "having processes in place for people to report abuse" provides you a space to "respond to it" is also a way of finding out how "your products are being misused". In this statement, we see that a feature (such as a reporting mechanism) can be used for multiple purposes, including the improvement of the system; if seen as part of the system. Another participant asks: "do we need agency and not just in our computing systems, but actually like in our democratic systems? In our, like, the various -isms that are in our governmental and political structures that we live in?" If we see technologies as part of our world, as we do in this toolkit, we can see how a participant can ask questions about agency not just in our computing systems, but in the world(s) we exist in - different systems of legality, politics, and policy impact how we interact with technologies (GDPR in the EU is perhaps a great example of this). The same participant asks an important rhetorical question: "Actually, what [...] if we change what we mean with the system, does [our conversation] have a completely different impact then?"

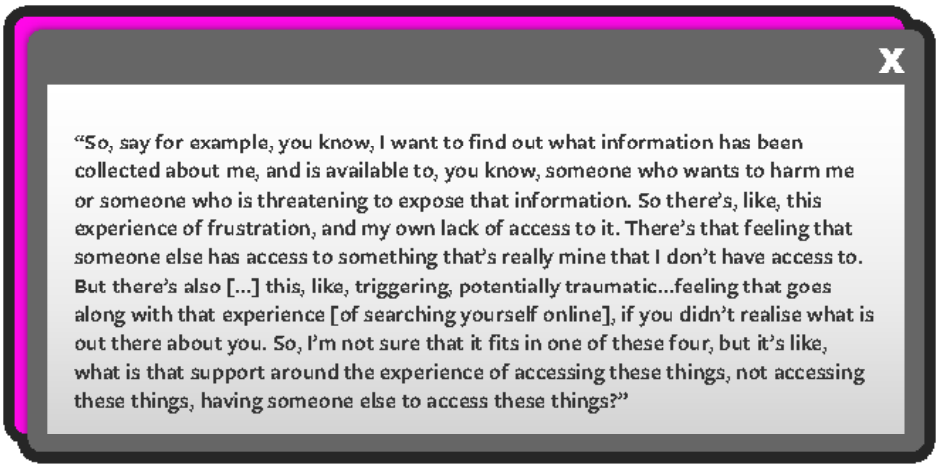
If we were to take a more holistic approach to HDI, we would need to not only speak to the interaction between people and their data on the individual level, but also on community and institutional levels. In doing so, we then must also take into consideration that people are individuals, that not everyone is the same; or as one participant said: "It just assumes that everybody is the same, or that everybody has the same abilities, the same rights, the same vocabularies and the same languages." In its current state, HDI seems like a totalizing framework that assumes everyone has the same degree of agency and authority to negotiate with those who use their data. This is not the case. Different people and the communities they are a part of will have different needs - and needs of people within a single community may also be different. For example, these needs relate to what 'legible' means to different communities. As one participant said: "To [one group of people] some of the [data] could be complete nonsense, almost like a different language, whereas to [...] different users it's understandable."

(2) There is a need for us to see and understand wider ecosystems in which HDI and support sit, rather than focusing only on the immediate issue

As mentioned above already, our technological issues can sometimes be mirrors to our wider social issues. However, it is important that we do not outsource our social issues to technology developers and designers; as one participant notes: "somewhere we have to acknowledge that we don't outsource our social issues to try to fix them with technology, because that, in all of these frameworks we always run the risk of doing this." For example, if software is developed for a company, workplace rules and hierarchies will constrain human agency as much as, if not more, than the software which implements those rules and hierarchies. This is an incredibly important point for us in critiquing the framework, but also in our extension of it. We do not advocate for "the outsourcing of our social issues or problems or concerns to technology designers as if they or we have those solutions, of like century long, millennia long human problems or human concerns."

Notions of legibility, agency, and negotiability can be useful, but they are not enough to ensure the ethical use of data. As one participant said, it is not enough to "put those principles out there without considering the kind of like underlying economic political processes that for why the data is being collected in the first place, and how that shapes what data we're collecting and what we're doing with it." This is because these terms do not address the issues that sit between these three tenets; or the things that underlie all of them: capitalism and patriarchy. As one participant said: the framework "kind of skirts around the issue of what the drive for the data collection is in the first place."

On top of skirting around the underlying issues of the data collection, the HDI framework also does not take into consideration the worries and anxiety that can bubble up in people if they are not aware what information is available about them online. This can be especially pronounced for people who are experiencing technology-mediated abuse and/or those who have experienced other forms of domestic violence, stalking, or other similar forms of violence. To summarise our second point of critique of the HDI framework, we ask how we can think about abuse that is perpetrated by people using data-intensive systems more holistically? How can we take into account the wider power structures that are at play? How can we ensure our data-intensive systems are built in a trauma-informed way? And how can we best support those who are experiencing abuse through these systems, and those who advocate for them?



(3) There is a need to be more process-oriented and less outcome-oriented

Thinking seriously about safety in relation to our data-intensive systems is not an easy task. One way that we can do this however, is to see the process of designing and developing new systems and/or services as equally, if not more important than the outcomes. This may seem contradictory to a design and technology industry that aims to ship new products regularly. But in reality, it is a reframing of how we already do our work. Updates and patches are shipped often, and can be seen as part of a software/hardware development process that is never quite 'finished' - there is always an update that can be pushed to make the system better. With this shift in thinking towards process, we can counter techno-solutionism and the arrogance that can come with this assumption that new technologies can solve our longstanding societal problems. For example, sometimes designers come into spaces of different expertise and create a new app or service that aims to solve a particularly complex problem that others have not been able to solve yet - or for which there may never be a solution.

Some technology companies and support organisations are already thinking in this way, as one participant notes: "I think there are so many conversations that are going on [...] and I really hope that, sort of, [technology companies] can speed themselves to be able to be there for these victims." As another participant points out though, even when data-intensive systems and the technology companies who build them provide support, "there's also something, like, a level of support that is missing." Survivors of technology-mediated abuse require more than responsive technology systems - they require adequate social, psychological, and other forms of support. This same participant goes on to make a point about the importance of funding in this space - pointing out that maybe it is a matter of "funding the people who provide this kind of support. So funding advocates...more...so that there can be more advocates, rather than, like, more technology solutions, or better technology."

To avoid techno-solutionism, designers and technologists must recognize the expertise of others - in the case of improving safety for people using data-intensive systems those people are survivors and advocates. When these people work together, equitably, with technologists, we can build and create safer technologies. And if we do that in a way that focuses on the process of this building, we can think about violence on our systems more meaningfully: we can address it at all stages of the design process, put in place feedback loops, and develop systems that support people: as one participant said, the "burden of safety really shouldn't fall solely on the shoulders of the end user. And that's why we felt is was important that at least some of the onus be shifted onto thoughtful design [...that helps...] technologists think about building their products from the ground up to be resistant to abuse." It is important to note though, that even if we work in the most equitable way, with the best intentions, we will not solve the societal issue of violence with new technologies.

Summary

In summary, we found that the HDI framework is useful in some ways, as it provides developers, designers, and technologists to think about how a person can engage with the processes of data collection, manipulation, and use by companies. However, it is hard to apply this logic to people who sit outside the box of the traditional 'user' that we think about when we design; even more so if that person has experienced or is currently experiencing the abuse of data or data-intensive systems to cause them physical, psychological, or other forms of harm. As one of the participants in our research noted, this is "maybe [because] the technical community is quite focussed on the optimistic side of their amazing technology that they're producing every day. That, it's difficult to think that something you produced, something you spent time writing, is being used to cause harm."

After reflecting on the HDI framework with researchers and practitioners, we concluded that it lacked a discrete discussion of power and that it does not situate data and data-intensive systems within wider ecosystems of use. Instead, it focuses solely on immediate issues at hand for individuals whose data is being used in ways that they do not approve of. We have already pointed towards some ways in which these issues can be addressed above, including a move towards being more process-oriented rather than output-oriented in the development of new data-intensive technologies. We believe that when we do this, we can learn to make small changes throughout a product development roadmap, with which we can make positive changes for those who are most vulnerable to technology-mediated abuse or other forms of exploitation and misuse of data-intensive systems. To be able to address this issue more meaningfully through the HDI framework, we present two new tenets with which we want to extend the existing HDI framework: trust and abusability.

A feminist extension of the Human-Data Interaction framework: Trust

ROSANNA BELLINI

Introduction

You would be hard pressed to find words as rich in meaning and application as the term trust. In most instances when we talk about trust, we talk about how trustworthy something is, such as "I trust this source of information because its from a reputable source"; or we may talk about how we place people in positions of responsibility, such as being in a position of trust as a developer; or perhaps even that we want to indicate that we expect something to happen, such as trusting that the software will act as planned. But what exactly are we talking about when we talk about trust, and, most importantly, why is it important to Human-Data Interaction? Trust is unique in that it has a special relationship with risk, with some scholars even arguing that only under conditions of risk is trust needed. Not all definitions will state this explicitly, with some describing trust through likelihood, confidence, and predictability

Curiously, trust is there even if we don't realise its presence. Marsh and Dibben (2003) argue, trust may underline every decision we make about what we do, how we do them and why. This means it also underlines our interactions with data-intensive systems and other technologies.

Maxine-Laurie Marshall (2018) expands this to claim that "trust is intrinsic to our everyday lives and the function of society as a whole". As we cannot predict for certain what will happen in our lives, we have to expose ourselves to the risk of uncertainty in our actions and beliefs, we have to trust in people, processes and situations. If we did not hold a degree of trust in anything, there is little in life we would be able to do. Yet given that trust is so pervasive in every interaction we have, it seems curious we have yet to explore what the implications for trust are, particularly when technology companies move at such a rapid pace that demand we place trust in something far faster than we would have been able to previously. This has significant implications for when we consider how such digital technologies may be leveraged to conduct interpersonal abuse. Notably the current zeitgeist has seen glimmers in works that question whether big technology companies should be permitted to hold such positions of trust over our personal information and data (Shipman & Marshall, 2020; Baig et al, 2020). However, we need to go further when examining how data, design and digital systems might be misused in cases of interpersonal violence to ensure that we do not inadvertently increase the negative experiences of already vulnerable people. Simply, if we are thinking about data, design, and digital systems we need to think about trust. And we need to think about trust not just at one point in the design cycle, but throughout the entire development and deployment cycle - we outline some ways this can be done below.

After conducting a rapid evidence review we examined how trust was being discussed in literature where a person of concern was actively misusing digital technologies (be it devices, systems, or services) to cause direct harm to another person. Based on our search criteria, we identified a range of different sources of information to look at on Human-Computer Interaction and Information Science; from technology-facilitated abuse in former and current intimate partnerships to organisational misuse of employee-tracking software. With the exception of a few notable works (Marques et al, 2019; Tseng et al, 2021), works did not explicitly position trust as an important factor to examining their research on interpersonal violence. In response to this lack of engagement with the terminology of 'trust', we formulated three factors that we deemed important to discuss further: the dynamism of interpersonal trust and encouraging researchers to explore meanings of 'trust' beyond the trustworthiness in systems. To end this section we highlight what developers, technologists, and researchers can do to interrogate the role 'trust' plays between the people who are intended to use their systems, the people who may be abusing the systems, and between the people who may be interacting through the system.

The dynamism of trust in interpersonal relationships

One of the most important qualities that trust possesses is the fact that it is dynamic; meaning that it can change based on external changes. Julia Slupska discusses how smart home threat models understand trust to be a static feature, when in actuality it can dramatically change over time (Slupska, 2019). Indeed, it is so volatile that many scholars describe it as a "weak-link concept" (Sherchan et al, 2013), that once a single violation of trust occurs (such as a data leak of personal information) it may be very hard or impossible to restore that level of trust to original levels. This dynamism makes this concept challenging to address, but at the same time absolutely crucial to work with in a time-sensitive and context-aware manner. It means that we need to get our understanding of what our dynamic and context-aware 'trust' looks like right the first time for there to be hope of delivering data-intensive services in ways that understand power, justice and equitable access. This is by no means an easy task as it is a very hard thing to get right, but understanding that trust looks different for different people in different contexts is a good starting point for challenging assumptions and in turn developing more useful and safe systems.

Looking beyond trustworthiness, and the need to understanding trust and its complexities

As mentioned previously, trust has a lot of different interpretations to different areas of knowledge and it is no surprise that many scholars have latched onto the psychological definition that prioritises scrutinising trustworthiness in people, sources of information, and systems. As Kittur et al. (2008) point out, arguing to increase trust in a system can equate to an overarching goal of encouraging people to use a system in the first place. In turn, this can have a positive impact such as increasing the depth of resources on Wikipedia. However, increasing engagement can be maliciously taken advantage of to feed the attention economy, ensuring that more people spend more time on a platform, and in some cases directly increasing the ways that people can spend money. Here, claims to increase trustworthiness, if used naively to be a replacement for recruitment into digital architecture where a user's attention, may lead to information and behaviour may be packed up for profit and sold to external companies (Zuboff, 2019). This more malicious intent however, is not always the conscious goals of the scholars or the companies who develop these services. Arguably, increasing the level of confidence and trust in information about one's health online is, at face value, apositive. However, in cases of interpersonal violence, such as image-based sexual abuse (McGlynn & Rackley, 2017; Powell et al, 2020), encouraging a victim-survivor of such abuse to trust in the reporting process of the same website that hosts material that causes them harm can further those feelings of harm by producing a trauma response. Indeed, we must question whether 'increasing the trustworthiness' of such situations can really bring about the outcomes desired by the end users; If we are genuinely interested in delivering justice, equality and fairness for such individuals, we must not simply define trust as a synonym for engagement and attention that can be exploited for profit, but seek to engage with what a positive belief in a process for justice might look like.

Asking the right questions

As Strohmayer has already discussed in a previous section in this toolkit, we argue that technology cannot simply provide the answers to problems without inadvertently creating new ones or exacerbating others. I build on this assertion that technologies can encourage us to examine new questions about ourselves and our connections with others; and we should have the confidence to ask questions of such devices. In our review, we kept returning to the same questions about trust, which we found useful to explore the relationship between technologies, trust, and abuse. These questions can be used to evaluate how data-intensive systems and the companies who own them talk about 'trust' to better understand why they want users to trust their system (ie. is it solely to improve engagement or a genuine attempt to improve safety of peoples using their system?) . These questions can also be used to interrogate new technologies that are being designed, but must be repeated several times throughout the design process and after deployment.

Is trust being used as a verb or a noun?

Going back to our school days for a moment, we may be reminded that a verb is a 'doing word' while a noun is a 'naming word'. The power behind these two descriptions is that of action, movement or, conversely, staticity. When we talk about trust as a noun, we have a tendency to understand it as somewhat fixed, permanent and its meaning shared and agreed upon by everyone involved. When we talk about trust as a verb however, we see the opposite: as flexible, temporal, and highly context-dependent. Its use in one way was not necessarily superior to another, as it holds different purposes depending on how it is used.

Who, where and why do we Trust?

This question targets exactly what we are talking about when we refer to contexts of trust. The **who** attempts to question if it is a person or an organisation that we are supposed to trust. Do we have assurances that a company will have its users in mind, or is it a case of placing trust in an individual at such a service; for example, someone managing a report of harm within the company? The **where** asks about the geographic or conceptual location of trust,such as a reporting process, a social innovation team, or a physical work space. Are we more likely to trust some of these contexts than others? And if this is the case, why is this the case? Finally the **why** asks about the purpose of trust in digital systems in a specific context. Why is it in our interests to trust a system or the company who runs it? Why does a company wish me to place my trust in it or a service it runs?

How do we Trust?

How we trust encourages us to examine the processes, manners, and ways in which we trust in digital systems. Research articles may not explore this as the answer may be implicit: using something means we trust in it. However, is it realistic to assume that someone has to trust in a process or system to use it in the first place? Or can we consider that people may distrust a service yet have no other choice than to engage with using it?

Summary

We hope the above questions help interrogate the meanings of trust in how, why, when, and where we use certain technologies. Looking towards the development of new systems, we have to continue to question what kinds of interpersonal trust we wish to support.

- Can your digital system really be trusted by victim-survivors of interpersonal violence mediated through the system?
- Do users of your system trust workers in your company enough to reach out to them when something harmful has happened?
- Can the people who use your system really trust in the reporting process and trust that using it will lead to an outcome that they are satisfied with?

While digital technologies haven't changed the definition of trust entirely, they have changed the process in how we trust in systems and in other people. Indeed, trust in a digital sense brings about new ways that we might trust what we expect of processes and other people, and how we might interact with each other around support and abuse. While this has brought about positive benefits such as trusting in caring services that try to mitigate technology-facilitated abuse, this can mean we trust too easily in processes that do not have our best interests at heart.

A feminist extension of the Human-Data Interaction framework: Abusability

JULIA SLUPSKA

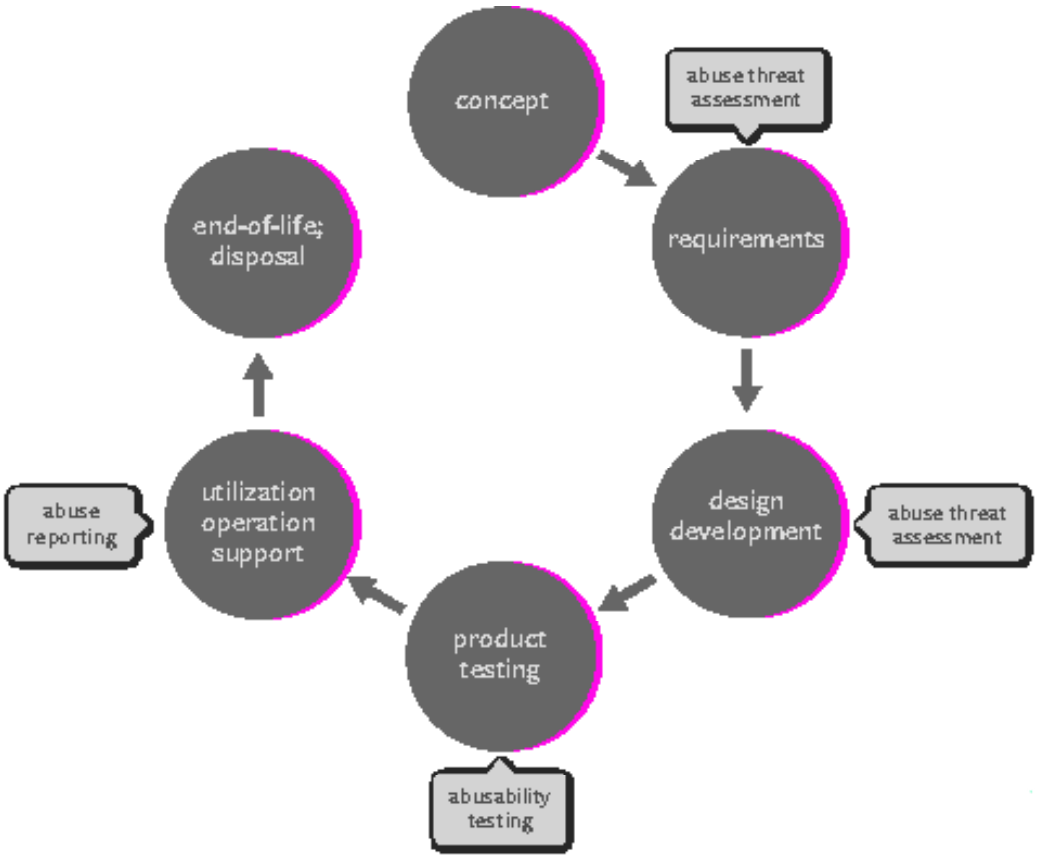
When developing new technologies, designers like to focus on the positive: what will this enable? What will this optimise? How will it change the world (for the better!)? As a discipline, information security often handles the negative, asking instead, how can something go wrong? Yet the way infosec experts ask these questions is usually too narrow, focusing on how an external malicious actor might penetrate a system to steal data or wreak havoc, rather than how an authorised user might (mis)use the system for harm. Likewise, designers tend to assume authorised users are not malicious. This section explores the concept of 'abusability', which offers a potential route to systematic and rigorous methods for mitigating abuse of technology, and why it is important to integrate this into the HDI framework.

Abusability is defined as the possibility that malicious actors might hijack or weaponise a system for harmful activity; designers have a responsibility to anticipate and mitigate this (Calderon et al, 2019). In this sense, abusability is similar to the concept of 'dual use scenarios' in cybersecurity and weapons research. Many technologies, like nuclear power or artificial intelligence, can be used for both military and peaceful purposes. The challenge of diplomacy on dual use technologies is to allow innovation while limiting harmful uses. Abusability addresses similar problems at the level of interpersonal, rather than international, relations.

The concept of "abusability" plays on the concept of "usability" to ask what kinds of uses should be restricted rather than enabled in design (Greenberg, 2019). In doing so, it invites designers to anticipate, mitigate and respond to abuse in the same way that designers might consider usability at various stages of the product development lifecycle.

Explaining the secure systems development lifecycle

The chart below depicts this process: the concept stage includes brainstorming, planning and market research. Engineers then take this concept and translating it into specific 'requirements' or features and functionality. At the design and development, developers implement these requirements in a model product. The model product is then tested for usability and other criteria, before being released. Once the product is on the market, maintenance involves utilisation and support for customers. Lastly, "end-of-life" involves disposal or recycling of the product, as well as the afterlife of the data created by the product of platform. Although this chart shows a linear progression, in reality products will go through multiple iterations, for example returning to the design or requirements stage for updates or if testing reveals some problems.



Abusability should be incorporated at various stages of the design lifecycle. As engineers set requirements (although this may also happen at the design stage), they should conduct an abuse threat modelling exercise asking: how might this product be abused for harm (Slupska and Tanczer, 2021; PenzeyMoog, 2021)? Once the product is developed, abusability tests include abuse scenarios akin to penetration testing, in which a malicious actor attempts to use the product for harm (Parkin et al, 2019). Results are fed back into product documentation or policies, or in more serious cases, the product is withdrawn or sent back for redesign. Once the product is on the market, the company offers a robust abuse reporting feature and support for rectification (Zou et al, 2020), monitoring which aspects of the product are being abused and looking for ways to mitigate this in the future.

Further resources for exploring abusability:

- Roxanne Leitao's work (2019) involving survivors of intimate partner violence in **anticipating harmful uses of smart home devices**
- IBM research (2020) on **Coercive Control Resistant Design** offers principles for designers seeking to address coercive control
- Chayn, an organisation which uses crowd sourcing to develop resources for survivors of domestic violence, has developed principles for **trauma-informed design** (2021)

Abuse of Power Comes as No Surprise

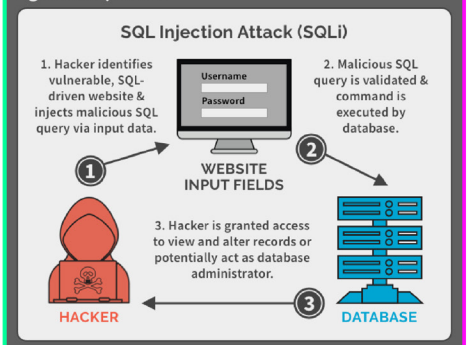
Abusability threat assessment and testing reframes security's traditional focus on an external attacker penetrating or subverting a system from its intended or authorised purpose, to ask how people use features for harm. In Soltani's initial formulation, these harms had to do with national security and international politics: for example, nation states using Facebook for election interference. However, we argue this concept is particularly helpful for thinking about technology-facilitated harm within intimate relationships.

The Cornell IPV Tech Team uses the concept of a "UI-bound adversary" (UI stands for user interface) to indicate an attacker who is not technologically sophisticated and therefore is limited to using a product's features and user interface in their attacks. For example, in an abusability attack, a perpetrator might co-opts the features of an app like 'Find My Friends' which is meant to help friends keep track of each others locations, to stalk or control the movements of a current or former partner. In contrast, a prototypical information security attack like an SQL injection, involves compromising a database by inputting malicious code. Both attacks involve subverting a system from its intended focus, but in the former, the perpetrator is using the features of the system "correctly" (but maliciously) while in the latter, the attacker is "breaking in" to the system. It is useful to remember that many typical abusability attacks do threaten classic information security principles like "confidentiality", "availability" or "integrity" and therefore these are not a separate classes of attack. However, abusability attacks are often missed or dismissed in infosec (information security) research (Slupska, 2019).

SECURITY VS ABUSABILITY:

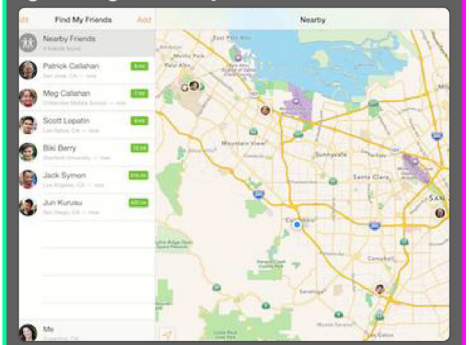
Prototypical Information Security Attack	Prototypical Abusability Attack
Compromising a system's confidentiality, integrity or availability through "breaking in"	Misusing a product's features for harms such as stalking or coercive control

Eg: SQL Injection



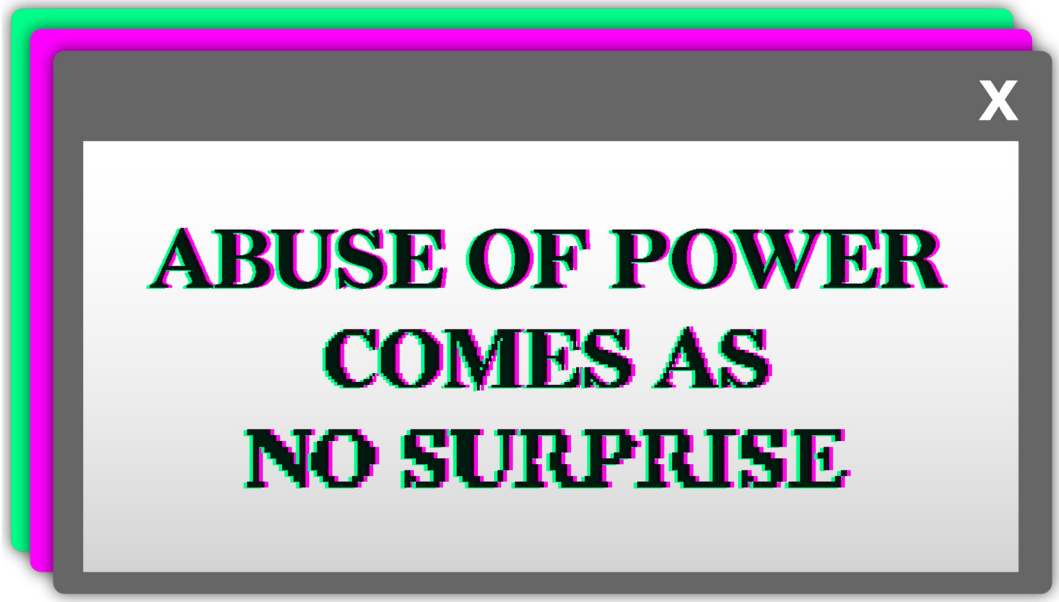
source: https://spanning.com/blog

Eg: Stalking via Find My Friends



source: https://en.wikipedia.org

How we understand what "counts" as a security issue, and what is worthy of designers and infosec experts' time, is a matter of how we understand power. By considering abusability through a feminist lens, we propose that technology designers must pay close attention to the power relations that are mediated by their products. Abusability is a reminder that power will often be abused, or as the old feminist slogan puts it **"abuse of power comes as no surprise."**



"Coercive control" is an act or a pattern of acts of assault, threats, humiliation and intimidation or other abuse that is used to harm, punish, or frighten. The term is used to make it clear that "domestic violence" does not always include physical violence, and does not only occur in cohabitating relationships. Coercive control has always happened in intimate relationships, and now these patterns of abuse are being reinvented with new technologies. Technology will not solve these problems, which are fundamentally a matter of values and power, but ignoring them only makes them worse.

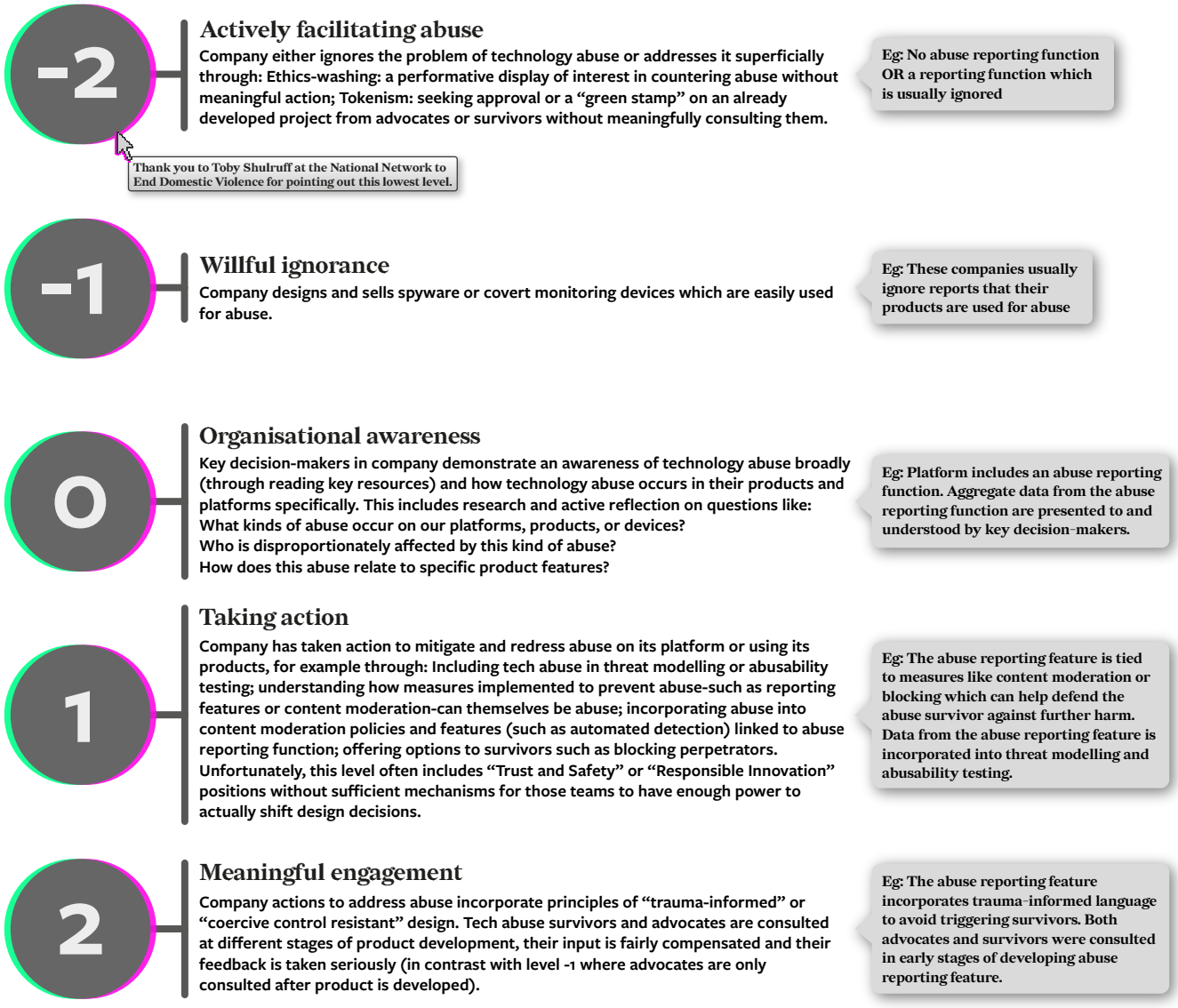
Building responsive systems

Although abusability is critical at the design stage (where it is preventative), it is also important to address abuse once it has happened. Both, the advocates who were interviewed for this project and those who attended our workshop, emphasized the importance of building response systems which have a relational dynamic with victim services: responding to problems as they arise and developing support systems for redressing cases of abuse.

The illustration below shows the different steps organisations can take to incorporate notions of abusability into their product design and organisational practices. Referencing the concept of maturity models, we present this as a development of maturity levels. We want to highlight that these steps relate to all sorts of forms of potential abuse of a system, but we are using a 'reporting abuse' feature as an example to illustrate the argument we are making. Adding to this, the companies who are perhaps most organizationally mature, might not be the companies who are most mature in relation to their understanding of abusability and their treatment of abuse on their platforms.

This diagram can help organisations mobilise to meaningfully incorporate abusability as concrete actions throughout their project lifecycles. This is a high level overview and not a full maturity model that would be used in industry for evaluating health and safety practices; however, such details could be fleshed out in the future.

MATURITY MODEL:



Summary

Through abusability, we focus on psychological, emotional and physical harm rather than damage to property or data in vulnerable systems (which is the focus of conventional information security) though these may also be involved. Instead of defending company profits against data breaches, we need to ask what groups and individuals are made marginal through abuse on our platforms and devices, and what kinds of harms are exacerbated by abuse.

Abusability plays on the concept of "usability" to ask what kinds of uses should be restricted rather than enabled in design. These questions should be incorporated at the design, testing and maintenance stage of product design. Mature company practices in this space will move beyond tokenistic inclusion and meaningfully involve survivors and advocates in designing both preventative measures and responses to abuse.

The concept of abusability is a critical addition to the HDI framework because it offers a practical way to incorporate power dynamics into designing human-data interactions. By understanding that agency is not always a good, as agency can also be a perpetrator's agency to harm others, we introduce more critical self-reflection into the design process. Feminists have long known that abuse of power comes as no surprise; it's time for tech companies to stop being surprised at this as well.

Case Studies:

As we have shown in the above extensions of the HDI framework, thinking about trust and abusability throughout a project lifecycle are incredibly important when wanting to design technologies that are safety-conscious. To better illustrate what the use of these terms can look like in practice, we present a series of three case studies below. In these case studies, we first describe a situation that has taken place, based on interviews with advocates for people who have experienced technology-mediated abuse. After this, we present a short analysis of this case study, followed by a series of questions that can be useful starting points for reflection.

We hope that these case studies can help kick-start conversations about trust and abusability. We also present them as resources at the end of the toolkit to be used as part of your meetings, trainings, teaching, or personal development.

CASE STUDY 1: COURT CASE WITH HIDDEN CAMERAS

A woman was recorded for 10 years in her home by her partner without her knowledge. Her partner is now using 10 years of security camera footage against her to fight a custody battle. The partner was able to self-select footage that suited his case and omitted evidence of his own behaviours. The court case was broadcast live due to covid-19. The advocate was able to support the woman by using her organisation's resources to help the woman "understand and validate her experience of coercive control, because sometimes she doubted her own experiences of whether she was in an abusive relationship" because the relationship had not been physically abusive. The advocate also encouraged her to collect evidence of threatening messages. "You start peeling the layers that society has, like you know, put on women's minds about compromise and understanding the other person and they start seeing the situation for what it is. I think that is a very heavily under- under appreciated like services to support survivors understanding"

Analysis

In this case, as in many cases of technology abuse, supporting survivors by validating and helping them articulate their experience of abuse as abuse is a critical part of the support advocates give

Questions for reflection

- How can platforms support survivors ability to collect evidence?
- How can platforms and legal systems prevent perpetrators from abusing mechanisms that are meant to ensure justice (like content reporting features or court cases)?

CASE STUDY 2: PORNHUB WEBSITE REFERRALS

Pornhub, without seeking or getting permission, links to an advocates' organisations' Facebook page on its "Non-Consensual Content Policy" website, which results in thousands of people from all over the world reaching out for support with cases of image-based sexual abuse. The advocate spends an increasing amount of her time helping people navigate Pornhub and other platforms, like Facebook's, non-consensual content policies. She said "the thing that's really disheartening and upsetting, is that, you know, someone reaches out to me to support them. Like immediately [...] like I'm really going to be like [...] Okay let me just get let me just get Mark on the phone quickly and I'm like yo Zuckerberg [...] take this down quickly."

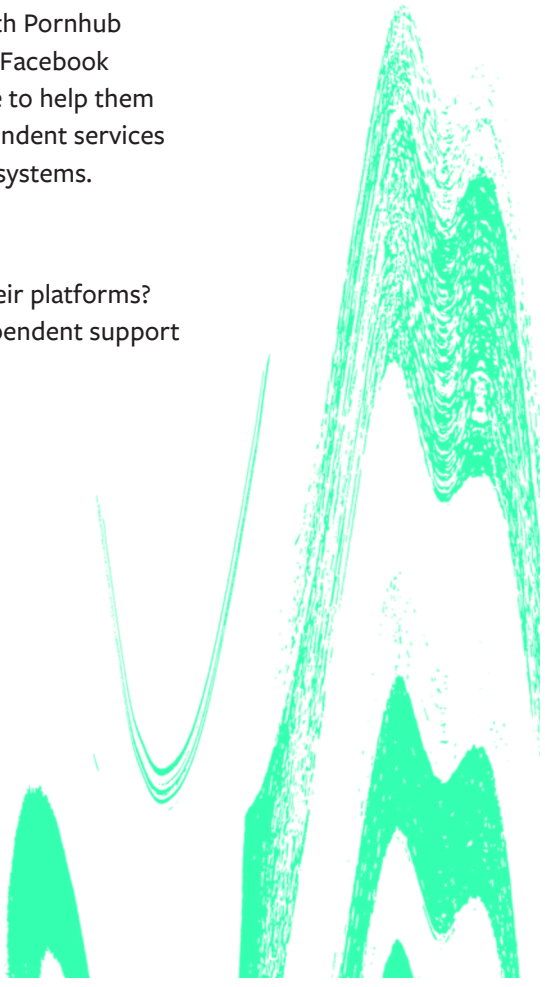
This is challenging as it often takes weeks to get non-consensual content removed, and then when you do get it removed, there is no support for getting evidence to prove it e.g. in a court of law. As a result, she said "It's like I don't have the funding anymore to do this work and I can't stop either right? [...] And it's not like- this isn't my role [..] I'm not a trained counsellor. But like I said, people just want to hear a soothing voice and you know somewhat be directed to what they need to do."

Analysis

Although it is good that Pornhub at least has a page for people who experience non-consensual content sharing, the fact that it is burdening these support services without compensating them for their labour is problematic. It would be better if both Pornhub contacted these agencies first. It would also be helpful if both Pornhub and Facebook offered some form of phone or email support to people experiencing abuse to help them with the process of takedowns. However, it is also invaluable to have independent services that provide emotional and technical support to survivors navigating these systems.

Questions for reflection

- What kind of support should companies provide survivors of abuse on their platforms?
- What are the pitfalls of companies providing this support instead of independent support services?



CASE STUDY 3: RING CAMERAS AND INTIMATE PARTNER VIOLENCE

Ring doorbell cameras are small cameras placed in peepholes which notify you via a smartphone or desktop app when anyone presses your doorbell. When installed by perpetrators or linked up to a perpetrators phone, Ring cameras can become an intrusive surveillance device, notifying an abuser about the survivors movements and guests to their home. This can isolate the survivor, as "now the perpetrator knows whenever she's leaving the house" (tech abuse advocate).

However, Ring cameras can also be used by survivors to secure their home against a stalker or an abusive ex-partner. They can help provide the survivor with a sense of security, or help them collect evidence of stalking used for a court case. For this reason, some advocates have reached out to Ring to ask for discounted cameras for survivors. One advocate described these doorbell cameras as "a real way for the client to kind of take back her safety and like a little bit of peace of mind."

Analysis

Ring cameras pose both risks and opportunities for survivors of intimate partner violence. They also bring survivors into broader systems of data collection and surveillance: in recent years, Ring (which is owned by Amazon) has partnered with hundreds of US law enforcement agencies, offering departments access to its platform so that police can request the video recorded by homeowners' cameras within a specific time and area (Harwell, 2019).

Questions for reflection

- Can one design a doorbell camera to be useful for survivors but not perpetrators of abuse?
- What are the implications of partnerships like Ring's partnership with police, particularly for marginalised communities?

SUMMARY

Across the three case studies, it is possible to see two important areas that appear time and time again: (1) the importance of evidence in cases where data-intensive technologies have been used by people to abuse others; and (2) the importance of recognizing expertise from non-technology sectors to improve data-intensive systems.

Evidence and abusability: Collecting evidence, for example, screenshots of abusive messages, or a data log which records account compromise, is critical for redressing abuse, especially through legal routes such as reporting to law enforcement, getting a restraining order, or going to court. However, we must remember that both evidence-collection and court systems are themselves subject to abuse.

Expertise: Advocates in domestic violence, sexual violence, and digital privacy support services are experts in safety and abuse through their work supporting survivors. Their insights on the ways that technology can be abused are critical for anticipating and mitigating abusability in technology design. Advocates also deserve to be compensated fairly for their insights. Unfortunately, technology companies can often treat advocates from civil society and social workers with a level of condescension. A common pitfall is that tech companies will develop a technology solution without advocate or survivor input, and then reaching out to advocacy groups only to test the pilot. Another pitfall is taking advocates time for granted, assuming they will be grateful for any tech company involvement at all.

References:

- Algorithms, V. C. F. L. and (2019) LogicLounge with Eva Galperin: Spouseware and Stalkerware - Where Do We Go From Here? @CAV2019 NYC, YouTube Video.

- Arief, B. et al. (2014) 'Sensible Privacy: How We Can Protect Domestic Violence Survivors Without Facilitating Misuse', in Proceedings of the 13th Workshop on Privacy in the Electronic Society. New York, NY, USA: ACM, pp. 201–204. doi: 10.1145/2665943.2665965.

- Australian eSafety Commissioner (2020) Children and technology-facilitated abuse in domestic family violence situations.

- Baig, K. et al. (2020) '"i'm hoping they're an ethical company that won't do anything that I'll regret": Users Perceptions of At-home DNA Testing Companies', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–13. doi: 10.1145/3313831.3376800. Barwulor, C. et al. (2020) '"Disadvantaged in the American-dominated internet":Sex, Work, and Technology'. SocArXiv. doi: 10.31235/OSF.IO/VZEHU.

- Bellini, R. et al. (2021) '"So-called privacy breeds evil" Narrative Justifications for Intimate Partner Surveillance in Online Forums', Proceedings of the ACM on Human-Computer Interaction. Association for Computing Machinery (ACM), 4(CSCW3), pp. 1–27. doi: 10.1145/3432909.

- Blunt, D. et al. (2020) Posting into the Void: Studying the Impact of Shadowbanning on Sex Workers and Activists.

- Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', Qualitative Research in Psychology, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

- Braun, V. and Clarke, V. (2019) 'Reflecting on reflexive thematic analysis', Qualitative Research in Sport, Exercise and Health. Routledge, pp. 589–597. doi: 10.1080/2159676X.2019.1628806.

- Braun, V. and Clarke, V. (2020) 'One size fits all? What counts as quality practice in (reflexive) thematic analysis?', Qualitative Research in Psychology. Routledge, pp. 1–25. doi: 10.1080/14780887.2020.1769238.

- Calderon, A. et al. (no date) AI Blindspot: A Discovery Process for preventing, detecting, and mitigating bias in AI systems. Available at: https://aiblindspot.media.mit.edu/ (Accessed: 15 October 2021).

- Chalmers, M. (2004a) 'A Historical View of Context', in Computer Supported Cooperative Work (CSCW). Available at: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/content/pdf/10.1007/s10606-004-2802-8.pdf&casa_token=jXKPxWngYbEAAAAA:WO-w-oU-wDsoFeaeIbQ_zQoZBGrQ5JlC4jGs67ahSoA72KxrMHrUvOcUWVvNybtKQGdm_Ef-w1xqdD3JYw (Accessed: 8 October 2021).

- Chalmers, M. (2004b) 'Hermeneutics, information and representation', European Journal of Information Systems. Palgrave Macmillan Ltd., 13(3), pp. 210–220. doi: 10.1057/palgrave.ejis.3000504.

- Chayn (no date) Trauma-informed design: understanding trauma and healing | by Hera Hussain | Chayn. Available at: https://blog.chayn.co/trauma-informed-design-understanding-trauma-and-healing-f289d281495c (Accessed: 15 October 2021).
- Citron, D. (2019) 'Sexual Privacy', The Yale Law Journal, 128(7).
- Coles-Kemp, L., Bjerg Jensen, R. and Heath, C. P. R. (2020) Too Much Information: Questioning Security in a Post-Digital Society, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. doi: https://doi.org/10.1145/3313831.3376214.
- Costanza-Chock, S. (2019) Design justice : community-led practices to build the worlds we need. The MIT Press.
- Crenshaw, K. (1989) 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics', University of Chicago Legal Forum, 1989. Available at: http://heinonline.org/HOL/Page?handle=hein.journals/uchclf1989&id=143&div=&collection=journals (Accessed: 10 May 2015).
- Dragiewicz, M. et al. (2018) 'Technology facilitated coercive control: domestic violence and the competing roles of digital media platforms', Feminist Media Studies. Routledge, 18(4), pp. 609–625. doi: 10.1080/14680777.2018.1447341.
- Dragiewicz, M. et al. (2019) 'Domestic violence and communication technology: Survivor experiences of intrusion, surveillance, and identity crime'. The Australian Communications Consumer Action Network (ACCAN). Eubanks, V. (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St St. Martin's Prss.
- Freed, D. et al. (2017) 'Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders', Proceedings of the ACM on Human-Computer Interaction. Association for Computing Machinery, 1(CSCW). doi: 10.1145/3134681.
- Freed, D. et al. (2018) '"A stalker's paradise": How intimate partner abusers exploit technology', in Conference on Human Factors in Computing Systems - Proceedings. Association for Computing Machinery. doi: 10.1145/3173574.3174241.
- Freed, D. et al. (2019) '"Is my phone hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence', Proceedings of the ACM on Human-Computer Interaction. Association for Computing Machinery, 3(CSCW). doi: 10.1145/3359304.
- Garritty, C. et al. (2021) 'Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews', Journal of Clinical Epidemiology. Elsevier Inc., 130, pp. 13–22. doi: 10.1016/j.jclinepi.2020.10.007.
- Glitch (2020) The Ripple Effect: Covid-19 and the epidemic of online abuse.
- Greenberg, A. (no date) Security Isn't Enough. Silicon Valley Needs 'Abusability' Testing | WIRED. Available at: https://www.wired.com/story/abusability-testing-ashkan-soltani/ (Accessed: 15 October 2021).
- Harris, B. A. and Woodlock, D. (2019) 'Digital Coercive Control: Insights From Two Landmark Domestic Violence Studies', The British Journal of Criminology. Oxford University Press, 59(3), pp. 530–550. doi: 10.1093/bjc/azy052.
- Harwell, D. (no date) Ring, the doorbell-camera firm, has partnered with 400 police forces, extending surveillance reach - The Washington Post. Available at: https://www.washingtonpost.com/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach/ (Accessed: 15 October 2021).

- Havron, S. et al. (2019) 'Clinical Computer Security for Victims of Intimate Partner Violence', in USENIX Conference on Security Symposium. Available at: https://www.usenix.org/conference/usenixsecurity19/presentation/havron.

- Hendricks, C. and Oliver, K. (1999) Language and liberation: feminism, philosophy, and language. State University of New York Press.

- Henry, N. and Flynn, A. (2019) 'Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support', Violence Against Women. SAGE Publications Inc., 25(16), pp. 1932–1955. doi: 10.1177/1077801219863881.

- Hörschelmann, K. and Reich, E. (2017) 'Entangled (In)Securities: Sketching the Scope of Geosocial Approaches for Understanding "Webs of (In)Security"1', Geopolitics. Routledge, 22(1), pp. 73–90. doi: 10.1080/14650045.2016.1214821.

- Kadri, T. E. (2020) 'Networks of Empathy', Utah Law Review, 2020. Available at: https://heinonline.org/HOL/Page?handle=hein.journals/utahlr2020&id=1113&div=33&collection=journals (Accessed: 15 October 2021).

- Kittur, A., Suh, B. and Chi, E. H. (2008) 'Can you ever trust a wiki? Impacting perceived trustworthiness in wikipedia', in Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW. New York, New York, USA: ACM Press, pp. 477–480. doi: 10.1145/1460563.1460639.

- Leitão, R. (2019) 'Anticipating smart home security and privacy threats with survivors of intimate partner abuse', in DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference. New York, NY, USA: Association for Computing Machinery, Inc, pp. 527–539. doi: 10.1145/3322276.3322366.

- Levy, J. and Jakobsson, P. (2014) 'Sweden's abolitionist discourse and law: Effects on the dynamics of Swedish sex work and on the lives of Sweden's sex workers', Criminology and Criminal Justice, 14(5), pp. 593–607. doi: 10.1177/1748895814528926.

- Levy, K. (2015) 'Intimate Surveillance', Idaho Law Review, 51(3).

- Levy, K. and Schneier, B. (2021) 'Privacy threats in intimate relationships', Journal of Cybersecurity. Oxford University Press, 6(1), pp. 1–13. doi: 10.1093/CYBSEC/TYAA006.

- Lopez-Neira, I. (no date) Gender and IoT | UCL Department of Science, Technology, Engineering and Public Policy - UCL – University College London, 2019. Available at: https://www.ucl.ac.uk/steapp/research/digital-technologies-policy-laboratory/gender-and-iot (Accessed: 15 October 2021).

- Marques, D. et al. (2019) 'Vulnerability & Blame: Making sense of unauthorized access to smartphones', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–13. doi: 10.1145/3290605.3300819.

- Marshall, M.-L. (no date) The dynamic nature of trust in the digital age - I-CIO | I-CIO. Available at: https://www.i-cio.com/big-thinkers/rachel-botsman/item/the-dynamic-nature-of-trust-in-the-digital-age (Accessed: 15 October 2021).

- Matthews, T. et al. (2017) 'Stories from survivors: Privacy & security practices when coping with intimate partner abuse', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 2189–2201. doi: 10.1145/3025453.3025875.

- McGlynn, C. and Rackley, E. (2017) 'Image-Based Sexual Abuse', Oxford Journal of Legal Studies. Oxford University Press, 37(3), pp. 534–561. doi: 10.1093/ojls/gqw033.

- Millar, K., Shires, J. and Tropina, T. (2021) Gender Approaches to Cybersecurity: Design, Defence and Response. Geneva, Switzerland. doi: 10.37559/GEN/21/01.
- Mortier, R. et al. (2013) Challenges & Opportunities in Human-Data Interaction. Cambridge.
- Mortier, R. et al. (2014) 'Human-Data Interaction: The Human Face of the Data-Driven Society', SSRN Electronic Journal. Elsevier BV. doi: 10.2139/ssrn.2508051.
- Mortier, R. et al. (2016) 'Human-Data Interaction', The Encyclopedia of Human-Computer Interaction. 2nd edn. Interaction Design Foundation. Available at: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-data-interaction (Accessed: 8 October 2021).
- Noble, S. U. (2018) Algorithms of oppression : how search engines reinforce racism.
- Nuttal, L. et al. (2019) Coercive Control Resistant Design. Available at: https://www.ibm.com/blogs/policy/wp-content/uploads/2020/05/CoerciveControlResistantDesign.pdf.
- Parkin, S. et al. (2019) 'Usability analysis of shared device ecosystem security: Informing support for survivors of IoT-facilitated tech-abuse', in ACM International Conference Proceeding Series. New York, NY, USA: Association for Computing Machinery, pp. 1–15. doi: 10.1145/3368860.3368861.
- PenzeyMoog, E. (2021) Design For Safety. A Book Apart.
- Plan International (2020) Free To Be Online? Girls' and young women's experiences of online harassment.
- Powell, A. et al. (2020) 'Image-based sexual abuse: An international study of victims and perpetrators – A Summary Report.' Royal Melbourne Institute of Technology.
- Project Shift | Creating a Safer Digital World for Young Women (no date). Available at: http://projectshift.ca/ (Accessed: 15 October 2021).
- Russell, L. (2020) Glitch Feminism: A Manifesto. Verso Books.
- Sambasivan, N. et al. (2019) '"They don't leave us alone anywhere we go": Gender and digital abuse in South Asia', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–14. doi: 10.1145/3290605.3300232.
- Sherchan, W., Nepal, S. and Paris, C. (2013) 'A survey of trust in social networks', ACM Computing Surveys. ACM PUB27 New York, NY, USA , 45(4), pp. 1–33. doi: 10.1145/2501654.2501661.
- Shipman, F. M. and Marshall, C. C. (2020) 'Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–12. doi: 10.1145/3313831.3376662.
- Sjoberg, L. (2018) 'Feminist security and security studies', in The Oxford Handbook of International Security. Oxford University Press, pp. 45–59. doi: 10.1093/oxfordhb/9780198777854.013.4.
- Slupska, J. (2019) 'Safe at Home: Towards a Feminist Critique of Cybersecurity', St. Anthony's International Review, (15).
- Slupska, J. and Tanczer, L. M. (2021) 'Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things', in The Emerald International Handbook of Technology Facilitated Violence and Abuse. Emerald Publishing Limited, pp. 663–688. doi: 10.1108/978-1-83982-848-520211049.

- Sugiura, L. and Smith, A. (2020) 'Victim Blaming, Responsibilization and Resilience in Online Sexual Abuse and Harassment', in Victimology. Springer International Publishing, pp. 45–79. doi: 10.1007/978-3-030-42288-2_3.

- Sultana, S. et al. (2021) '"Unmochon": A tool to combat online sexual harassment over facebook messenger', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–18. doi: 10.1145/3411764.3445154.

- Thomas, K. et al. (2021) 'SoK: Hate, harassment, and the changing landscape of online abuse', in Proceedings - IEEE Symposium on Security and Privacy. Institute of Electrical and Electronics Engineers Inc., pp. 247–267. doi: 10.1109/SP40001.2021.00028.

- Tseng, E. et al. (2020) 'The tools and tactics used in intimate partner surveillance: an analysis of online infidelity forums', in USENIX Conference on Security Symposium. ACM. Available at: https://dl.acm.org/doi/10.5555/3489212.3489319.

- Tseng, E. et al. (2021) 'A digital safety dilemma: Analysis of computer-mediated computer security interventions for intimate partner violence during covid-19', in Conference on Human Factors in Computing Systems - Proceedings. New York, NY, USA: Association for Computing Machinery, pp. 1–17. doi: 10.1145/3411764.3445589.

- Victorelli, E. Z. et al. (2020) 'Understanding human-data interaction: Literature review and recommendations for design', International Journal of Human Computer Studies. Academic Press, 134, pp. 13–32. doi: 10.1016/j.ijhcs.2019.09.004.

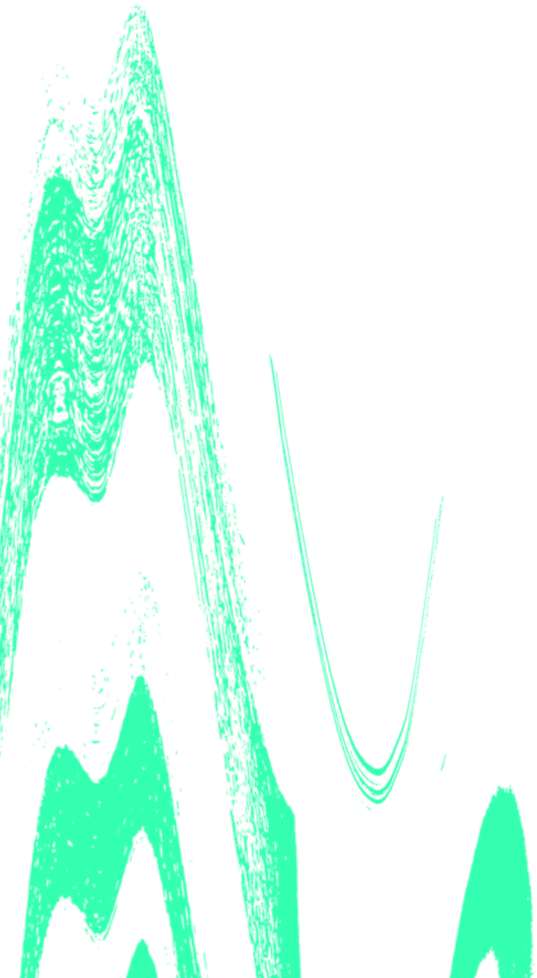
- Wajcman, J. (2007) 'From women and technology to gendered technoscience', in Information Communication and Society. Routledge , pp. 287–298. doi: 10.1080/13691180701409770.

- WESNET (2020) Second National Survey on Technoogy Abuse and Domestic Violence.

- Woodlock, D. (2017) 'The Abuse of Technology in Domestic Violence and Stalking', Violence Against Women. SAGE Publications Inc., 23(5), pp. 584–602. doi: 10.1177/1077801216646277.

- Zou, Y. et al. (2021) 'The Role of Computer Security Customer Support in Helping Survivors of Intimate Partner Violence', in USENIX Conference on Security Symposium.

- Zuboff, S. (2019) The age of surveillance capitalism : the fight for the future at the new frontier of power.



Resources:

In this final section of the Feminist HDI toolkit, we present a series of resources that bring together many of the points we have raised in Parts 1 and 2. These resources are stand-alone documents that can be printed separately, photocopied, hung up on a wall, and shared offline.

These resources include ideas for the design of new technologies to ensure they take trust and abusability into consideration. This also includes implications for technology companies who want to protect those who use their systems - we provide ideas how companies can better support survivors and treat perpetrators of online harms.

Following this, we take a look at the design and development cycles of new systems. We present the 'secure system engineering life cycle' as an illustration that can be shared with designers in trainings and in a format where this can be hung up on an office wall. This cycle is also illustrated with an example that presents different stages at which reporting functions can be used to promote safer engagement with data-intensive systems.

Finally, we point towards a different audience: journalists. We present a series of good practice guidelines that will help journalists write about harms, abusability, and trust in data-intensive systems.

We present 5 resources:

- Implications for advocates and others who support survivors
- Implications for designing technologies for survivors and perpetrators
- Describing the secure systems development lifecycle
  Case Studies: examples, analysis, and questions for discussion
- Taking Abusability seriously in your technology company or system: Presenting levels of maturity
- Do's and Don'ts for journalists covering technology-mediated abuse

We hope that the flyers, posters, and infographics will be useful for people in working towards developing safer data-intensive systems and technologies.

IMPLICATIONS FOR ADVOCATES AND OTHERS WHO SUPPORT SURVIVORS

Julia Slupska

This advice is drawn from interviews with other advocates working on digital privacy and technology-facilitated abuse, as well as our review of background literature.

RECOMENDATIONS FOR SUPPORTING SURVIVORS:

- **Maintain a general awareness of technology abuse and how to identify it**, as well as general digital security precautions like password management, multifactor authentication, and VPNs. This is not meant to be an introductory resource for learning about technology abuse: for that we recommend Refuge's Tech Safety site (https://refugetechsafety.org/).
- **Explaining how to download and use these tools step by step**: instead of giving general advice, "don't be afraid to get your hands a little dirty when it comes to technology and give actionable advice whenever possible"
- **Be confident in what you do know**: a lack of confidence in engaging with technology can be an unnecessary limitation. Many advocates repeated how often "just googling it" can prove to be a significant help to survivors.
- **Pay attention to children's accounts and devices** in threat assessment
- **Be realistic with threat models**: a particular risk for advocates coming from the digital privacy or cybersecurity space is "judging [survivors] very harshly, scaring them, giving them advice that is meant for protecting them from nation states or law enforcement rather than their [...] abusers level of technical skill."
- **Avoid telling survivors to get off social media** as that can be detrimental to survivors.

RECOMENDATIONS FOR ENGAGING WITH TECH COMPANIES:

- **Remember you have valuable expertise in technology safety and abuse:** knowledge of how things go wrong on the ground is critical and should be respected by tech designers.
- **Don't be afraid to ask for compensation for consultation:** tech companies should respect and compensate your expertise; it is inappropriate to ask you to contribute these for free.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

DESIGNING FOR SURVIVORS AND PERPETRATORS:

Rosie Bellini & Julia Slupska

When designing to mitigate abusability, you are necessarily designing for both survivors and perpetrators. This sentence may not initially make sense, and you may be asking; why would I want to design for perpetrators? If we are to consider how to make their products and/or services more safe, welcoming, and empowering for survivors, we must also think about how to prevent or limit abusive behaviours. While this may be complicated and challenging, it is worth the extra consideration both to protect marginalised users and because, as Penzey-Moog (2021) argues, designing for survivors makes products and services better for everyone (or at least, everyone who is not trying to cause harm). For example, taking an abusive use case into consideration leads to features like separating shared accounts that might help anyone going through other life situations, such as a break-up.

In this section, we outline some considerations for designing for survivors and perpetrators. We look at dating apps as a case study due to their close proximity to intimate partner violence and sexual violence.

SURVIVORS

Implication	Case study: dating apps
CONSENT users should have as much control as possible over how others can interact with them and how their information is shared	One positive example is an app that allows users to limit sharing of sensitive details, such as age, location, or sexual preferences. Another way an app can be consent-oriented is by allowing users to restrict who can contact them, for example to only allow 'matches' which the user has selected to message them.
RETRAUMATISATION designers should consider how certain things can be triggering to people who have experienced trauma	A feature which continually shares other users' proximity (for example, disclosing they are 5 kilometers away) can create a lot of anxiety for someone who has experienced stalking or is currently in an abusive relationship, reminding them that their perpetrator is still nearby. If this feature exists, it should be possible and easy to turn it off.
RESTITUTION approaches drawing on restorative justice focus on meeting the needs of a person who has been harmed	If users are exposed to abuse, users are quickly able to smartly evidence this abuse and attach other offending behaviours to a blocking system that is transparent, encouraging and extends support beyond the original blocking request. When responding to complaints, platforms should try as much as possible to meet the needs of the person who experienced harm, for example by keeping survivors informed on how their complaints are addressed (i.e. will the perpetrator be removed from the platform, why or why not) to provide closure, rather than leaving them wondering whether their complaint led to anything.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

PERPETRATORS

Implication	Case study: dating apps
PREVENTATIVE ACTIONS/MESSAGING users who are writing unacceptable messages could be alerted to this	If an abusive phrase or slur is detected in the system, it alerts the sender via a notification that this language is unacceptable on the dating platform. A smart messaging system may suggest an alternative or prevent the message from being sent. If an alternative is unavailable, for example, if a potential perpetrator is struggling with desisting from using abusive language, a notification/alert is displayed to the perpetrator of the community guidelines for using the platform.
ANTICIPATING FORMS OF ABUSE when designers know about risky features, these should be opt-in	Dating app companies are aware that their location systems may be used for stalking, and as such they turn off location tracking by default for all users. If a person's location is being continuously monitored then a system notification may request why this potential perpetrator is behaving in this way.
ACCOUNTABILITY systems should explain to people their reasons for being removed from the system to encourage accountability rather than simply punishing	Simply blocking content or removing a user from a platform does not really encourage accountability and may even discourage the perpetrator from taking responsibility (Salehi, 2020). Features that explain to a user why their behaviour is harmful encourage perpetrators to understand harm and work to repair it, for example, by working to avoid the behaviour in the future or even sending an apology. However, contact from the perpetrator may not be appropriate for many abuse cases, so this should be carefully mediated and only permitted with the survivor's consent.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

SUGGESTIONS FOR HOW YOU COULD PICK A FEATURE AND DO THIS YOURSELVES

First things first, **identify a feature**. A feature is a unique aspect that allows you to do something with a system, service or device. For example, a messaging app might have several features around communication, such as:

- Read receipts
- Share animated stickers
- Disappearing messages

Then, identify what the initial goal of the introduction of that feature was, asking what did a company or provider want their users to achieve with this feature?

READ RECEIPTS	An indicator that lets the sender know the receiver has a) received their message and/or b) read the message
SHARE ANIMATED STICKERS	A way of expanding communication for the sender beyond text such as emotions or reactions
DISAPPEARING MESSAGES	A way of keeping a conversation private between senders and receivers.

This helps to identify where the power imbalance might be, such as who is in the driving seat of these features (not forgetting the designers and developers who implemented them of course) and consider how such a feature could be used to cause harm. This is the tricky part of the exercise, but consider placing yourself in the shoes of a user who has just had a bad interaction with a feature and may come away feeling frustrated.

READ RECEIPTS	An indicator that lets the sender know the receiver has a) received their message and/or b) read the message	A sender could harass a receiver as to why they have not replied to their messages.
SHARE ANIMATED STICKERS	A way of expanding communication for the sender beyond text such as emotions or reactions	A sender could share intimidating or 'coded' messages through imagery.
DISAPPEARING MESSAGES	A way of keeping a conversation private between senders and receivers.	A sender could send abusive texts and then rely on the disappearing message feature to remove the evidence.

Here we are designing for perpetrators, a hypothetical perpetrator who may cause damage and harm. We also should consider that if someone is using bad behaviours, we could consider pairing these features with notifications that provide support or services for abusive behaviours in the local area.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

Designing for Survivors a

DATING APPS AS A CASE STUDY

Rosie Bellini & Julia Slupska

When designing to mitigate abusability, you are necessarily designing for both survivors and perpetrators. This sentence m
If we are to consider how to make their products and/or services more safe, welcoming, and empowering for survivors an
is worth the extra consideration both to protect marginalised users and because, as Penzey-Moog (2021) argues, designing
trying to cause harm). For example, taking an abusive use case into consideration leads to features like separating shared a
section, we outline some considerations for designing for survivors and perpetrators. We look at dating apps as a case stu

SURVIVORS

Implication	Case study: dating apps
CONSENT users should have as much control as possible over how others can interact with them and how their information is shared	One positive example is an app that allows users to limit sharing of sensitive details, such as age, location, or sexual preferences. Another way an app can be consent-oriented is by allowing users to restrict who can contact them, for example to only allow 'matches' which the user has selected to message them.
RETRAUMATISATION designers should consider how certain things can be triggering to people who have experienced trauma	A feature which continually shares other users' proximity (for example, disclosing they are 5 kilometers away) can create a lot of anxiety for someone who has experienced stalking or is currently in an abusive relationship, reminding them that their perpetrator is still nearby. If this feature exists, it should be possible and easy to turn it off.
RESTITUTION approaches drawing on restorative justice focus on meeting the needs of a person who has been harmed	If users are exposed to abuse, users are quickly able to smartly evidence this abuse and attach other offending behaviours to a blocking system that is transparent, encouraging and extends support beyond the original blocking request. When responding to complaints, platforms should try as much as possible to meet the needs of the person who experienced harm, for example by keeping survivors informed on how their complaints are addressed (i.e. will the perpetrator be removed from the platform, why or why not) to provide closure, rather than leaving them wondering whether their complaint led to anything.

PERPETRATORS

Implication	Case study: dating apps
PREVENTATIVE ACTIONS/MESSAGING users who are writing unacceptable messages could be alerted to this	If an abusive phrase or slur is detected in the system, it alerts the sender via a notification that this language is unacceptable on the dating platform. A smart messaging system may suggest an alternative or prevent the message from being sent. If an alternative is unavailable, for example, if a potential perpetrator is struggling with desisting from using abusive language, a notification/alert is displayed to the perpetrator of the community guidelines for using the platform.
ANTICIPATING FORMS OF ABUSE when designers know about risky features, these should be opt-in	Dating app companies are aware that their location systems may be used for stalking, and as such they turn off location tracking by default for all users. If a person's location is being continuously monitored then a system notification may request why this potential perpetrator is behaving in this way.
ACCOUNTABILITY systems should explain to people their reasons for being removed from the system to encourage accountability rather than simply punishing	Simply blocking content or removing a user from a platform does not really encourage accountability and may even discourage the perpetrator from taking responsibility (Salehi, 2020). Features that explain to a user why their behaviour is harmful encourage perpetrators to understand harm and work to repair it, for example, by working to avoid the behaviour in the future or even sending an apology. However, contact from the perpetrator may not be appropriate for many abuse cases, so this should be carefully mediated and only permitted with the survivor's consent.

SUGG

First thin
service o

- Read
- Share
- Disapp

Then, ide
what did

READ
SHARE
DISAPP

This help
forgettin
be used
who has

READ
SHARE
DISAPP

Here we
consider
provide s

nd Perpetrators:

nay not initially make sense, and you may be asking; why would I want to design for perpetrators? d how to prevent or limit abusive behaviours. While this may be complicated and challenging, it g for survivors makes products and services better for everyone (or at least, everyone who is not accounts that might help anyone going through other life situations, such as a break-up. In this dy due to their close proximity to intimate partner violence and sexual violence.

ESTIONS FOR HOW YOU COULD PICK A FEATURE AND DO THIS YOURSELVES

gs first, **identify a feature**. A feature is a unique aspect that allows you to do something with a system, r device. For example, a messaging app might have several features around communication, such as:

receipts
animated stickers
pearing messages

ntify what the initial **goal** of the introduction of that feature was, asking a company or provider want their users to achieve with this feature?

RECEIPTS	An indicator that lets the sender know the receiver has a) received their message and/or b) read the message
ANIMATED STICKERS	A way of expanding communication for the sender beyond text such as emotions or reactions
PEARING MESSAGES	A way of keeping a conversation private between senders and receivers.

ps to identify where the power imbalance might be, such as who is in the driving seat of these features (not ng the designers and developers who implemented them of course) and consider how such a feature could to cause harm. This is the tricky part of the exercise, but consider placing yourself in the shoes of a user just had a bad interaction with a feature and may come away feeling frustrated.

RECEIPTS	An indicator that lets the sender know the receiver has a) received their message and/or b) read the message	A sender could harass a receiver as to why they have not replied to their messages.
ANIMATED STICKERS	A way of expanding communication for the sender beyond text such as emotions or reactions	A sender could share intimidating or 'coded' messages through imagery.
PEARING MESSAGES	A way of keeping a conversation private between senders and receivers.	A sender could send abusive texts and then rely on the disappearing message feature to remove the evidence.

are designing for perpetrators, a hypothetical perpetrator who may cause damage and harm. We also should that if someone is using bad behaviours, we could consider pairing these features with notifications that support or services for abusive behaviours in the local area.

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

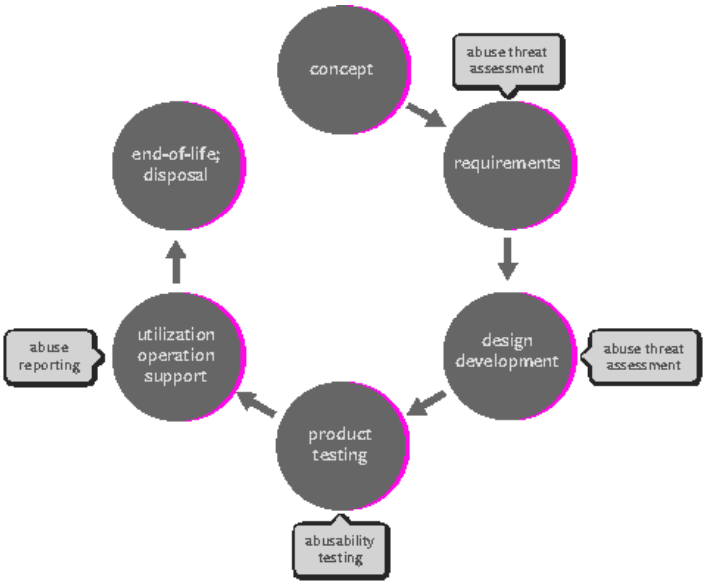
EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

ABUSABILITY AND THE SECURE SYSTEMS DEVELOPMENT LIFECYCLE

Julia Slupska

The chart below depicts the secure systems development lifecycle: the concept stage includes brainstorming, planning and market research. Engineers then take this concept and translate it into specific 'requirements' or features and functionality. At the design and development, developers implement these requirements in a model product. The model product is then tested for usability and other criteria, before being released. Once the product is on the market, maintenance involves utilisation and support for customers. Lastly, "end-of-life" involves disposal or recycling of the product. Although this chart shows a linear progression, in reality products will go through multiple iterations, for example returning to the design or requirements stage if testing reveals some problems.

Abusability should be incorporated at various stages of the design lifecycle. As engineers, set requirements (although this may also happen at the design stage), they should conduct an abuse threat modelling exercise asking: how might this product be abused for harm (see Slupska and Tanczer 2021, PenzeyMoog 2021)?



Abusability should be incorporated at various stages of the design lifecycle. As engineers set requirements (although this may also happen at the design stage), they should conduct an abuse threat modelling exercise asking: how might this product be abused for harm (Slupska and Tanczer, 2021; PenzeyMoog, 2021)? Once the product is developed, abusability tests include abuse scenarios akin to penetration testing, in which a malicious actor attempts to use the product for harm (Parkin et al, 2019). Results are fed back into product documentation or policies, or in more serious cases, the product is withdrawn or sent back for redesign. Once the product is on the market, the company offers a robust abuse reporting feature and support for rectification (see Zou et al 2020), monitoring which aspects of the product are being abused and looking for ways to mitigate this in the future.

FURTHER RESOURCES

- Roxanne Leitao's work (2019) involving survivors of intimate partner violence in anticipating harmful uses of smart home devices
- IBM research (2020) on Coercive Control Resistant Design offers principles for designers seeking to address coercive control
- Chayn, an organisation which uses crowd sourcing to develop resources for survivors of domestic violence, has developed principles for trauma-informed design (2021)

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

COURT CASE WITH HIDDEN CAMERAS

Case Study 1

A woman was recorded for 10 years in her home by her partner without her knowledge. Her partner is now using 10 years of security camera footage against her to fight a custody battle. The partner was able to self-select footage that suited his case and omitted evidence of his own behaviours. The court case was broadcast live due to covid-19. The advocate was able to support the woman by using her organisation's resources to help the woman "understand and validate her experience of coercive control, because sometimes she doubted her own experiences of whether she was in an abusive relationship" because the relationship had not been physically abusive. The advocate also encouraged her to collect evidence of threatening messages. "You start peeling the layers that society has, like you know, put on women's minds about compromise and understanding the other person and they start seeing the situation for what it is. I think that is a very heavily under- under appreciated like services to support survivors understanding".

ANALYSIS

In this case, as in many cases of technology abuse, supporting survivors by validating and helping them articulate their experience of abuse as abuse is a critical part of the support advocates give.

QUESTIONS FOR REFLECTION

- How can platforms support survivors ability to collect evidence?
- How can platforms and legal systems prevent perpetrators from abusing mechanisms that are meant to ensure justice (like content reporting features or court cases)?

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

PORNHUB WEBSITE REFERRALS

Case Study 2

Pornhub, without seeking or getting permission, links to an advocates' organisations' Facebook page on its "Non-Consensual Content Policy" website, which results in thousands of people from all over the world reaching out for support with cases of image-based sexual abuse. The advocate spends an increasing amount of her time helping people navigate Pornhub and other platforms, like Facebook's, non-consensual content policies. She said "the thing that's really disheartening and upsetting, is that, you know, someone reaches out to me to support them. Like immediately [...] like I'm really going to be like [...] Okay let me just get let me just get Mark on the phone quickly and I'm like yo Zuckerberg [...] take this down quickly."

This is challenging as it often takes weeks to get non-consensual content removed, and then when you do get it removed, there is no support for getting evidence to prove it e.g. in a court of law. As a result, she said "It's like I don't have the funding anymore to do this work and I can't stop either right? [...] And it's not like- this isn't my role [..] I'm not a trained counsellor. But like I said, people just want to hear a soothing voice and you know somewhat be directed to what they need to do."

ANALYSIS

Although it is good that Pornhub at least has a page for people who experience non-consensual content sharing, the fact that it is burdening these support services without compensating them for their labour is problematic. It would be better if both Pornhub contacted these agencies first. It would also be helpful if both Pornhub and Facebook offered some form of phone or email support to people experiencing abuse to help them with the process of takedowns. However, it is also invaluable to have independent services that provide emotional and technical support to survivors navigating these systems.

QUESTIONS FOR REFLECTION

- What kind of support should companies provide survivors of abuse on their platforms?
- What are the pitfalls of companies providing this support instead of independent support services?

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

RING CAMERAS AND INTIMATE PARTNER VIOLENCE

Case Study 3

Ring doorbell cameras are small cameras placed in peepholes which notify you via a smartphone or desktop app when anyone presses your doorbell. When installed by perpetrators or linked up to a perpetrators phone, Ring cameras can become an intrusive surveillance device, notifying an abuser about the survivors movements and guests to their home. This can isolate the survivor, as "now the perpetrator knows whenever she's leaving the house" (tech abuse advocate).

However, Ring cameras can also be used by survivors to secure their home against a stalker or an abusive ex-partner. They can help provide the survivor with a sense of security, or help them collect evidence of stalking used for a court case. For this reason, some advocates have reached out to Ring to ask for discounted cameras for survivors. One advocate described these doorbell cameras as "a real way for the client to kind of take back her safety and like a little bit of peace of mind."

ANALYSIS

Ring cameras pose both risks and opportunities for survivors of intimate partner violence. They also bring survivors into broader systems of data collection and surveillance: in recent years, Ring (which is owned by Amazon) has partnered with hundreds of US law enforcement agencies, offering departments access to its platform so that police can request the video recorded by homeowners' cameras within a specific time and area (Harwell 2019).

QUESTIONS FOR REFLECTION

- Can one design a doorbell camera to be useful for survivors but not perpetrators of abuse?
- What are the implications of partnerships like Ring's partnership with police, particularly for marginalised communities?

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

Case Studies

COURT CASE WITH HIDDEN CAMERAS

A woman was recorded for 10 years in her home by her partner without her knowledge. Her partner is now using 10 years of security camera footage against her to fight a custody battle. The partner was able to self-select footage that suited his case and omitted evidence of his own behaviours. The court case was broadcast live due to covid-19. The advocate was able to support the woman by using her organisation's resources to help the woman "understand and validate her experience of coercive control, because sometimes she doubted her own experiences of whether she was in an abusive relationship" because the relationship had not been physically abusive. The advocate also encouraged her to collect evidence of threatening messages. "You start peeling the layers that society has, like you know, put on women's minds about compromise and understanding the other person and they start seeing the situation for what it is. I think that is a very heavily under- under appreciated like services to support survivors understanding"

ANALYSIS

In this case, as in many cases of technology abuse, supporting survivors by validating and helping them articulate their experience of abuse as abuse is a critical part of the support advocates give.

QUESTIONS FOR REFLECTION

- How can platforms support survivors ability to collect evidence?
- How can platforms and legal systems prevent perpetrators from abusing mechanisms that are meant to ensure justice (like content reporting features or court cases)?

PORNHUB WEBS REFERRALS

Pornhub, without seeking or getting per organisations' Facebook page on its "No website, which results in thousands of p reaching out for support with cases of i advocate spends an increasing amount Pornhub and other platforms, like Facel policies. She said "the thing that's really you know, someone reaches out to me like I'm really going to be like [...] Okay the phone quickly and I'm like yo Zucker

This is challenging as it often takes weel removed, and then when you do get it r getting evidence to prove it e.g. in a cou I don't have the funding anymore to do [...] And it's not like- this isn't my role [... I said, people just want to hear a soothi directed to what they need to do."

ANALYSIS

Although it is good that Pornhub at leas experience non-consensual content sha these support services without compen problematic. It would be better if both P first. It would also be helpful if both Por form of phone or email support to peop with the process of takedowns. Howeve independent services that provide emot survivors navigating these systems.

QUESTIONS FOR REFLECTION

- What kind of support should compan their platforms?
- What are the pitfalls of companies pr independent support services?

SITE

rmission, links to an advocates' on-Consensual Content Policy" people from all over the world mage-based sexual abuse. The of her time helping people navigate book's, non-consensual content disheartening and upsetting, is that, to support them. Like immediately [...] et me just get let me just get Mark on rberg [...] take this down quickly."

ks to get non-consensual content emoved, there is no support for rt of law. As a result, she said "It's like this work and I can't stop either right?] I'm not a trained counsellor. But like ng voice and you know somewhat be

t has a page for people who ring, the fact that it is burdening sating them for their labour is Pornhub contacted these agencies rnhub and Facebook offered some ple experiencing abuse to help them r, it is also invaluable to have tional and technical support to

ies provide survivors of abuse on

oviding this support instead of

RING CAMERAS AND INTIMATE PARTNER VIOLENCE

Ring doorbell cameras are small cameras placed in peepholes which notify you via a smartphone or desktop app when anyone presses your doorbell. When installed by perpetrators or linked up to a perpetrators phone, Ring cameras can become an intrusive surveillance device, notifying an abuser about the survivors movements and guests to their home. This can isolate the survivor, as "now the perpetrator knows whenever she's leaving the house" (tech abuse advocate).

However, Ring cameras can also be used by survivors to secure their home against a stalker or an abusive ex-partner. They can help provide the survivor with a sense of security, or help them collect evidence of stalking used for a court case. For this reason, some advocates have reached out to Ring to ask for discounted cameras for survivors. One advocate described these doorbell cameras as ""a real way for the client to kind of take back her safety and like a little bit of peace of mind."

ANALYSIS

Ring cameras pose both risks and opportunities for survivors of intimate partner violence. They also bring survivors into broader systems of data collection and surveillance: in recent years, Ring (which is owned by Amazon) has partnered with hundreds of US law enforcement agencies, offering departments access to its platform so that police can request the video recorded by homeowners' cameras within a specific time and area (Harwell 2019).

QUESTIONS FOR REFLECTION

- Can one design a doorbell camera to be useful for survivors but not perpetrators of abuse?
- What are the implications of partnerships like Ring's partnership with police, particularly for marginalised communities?

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

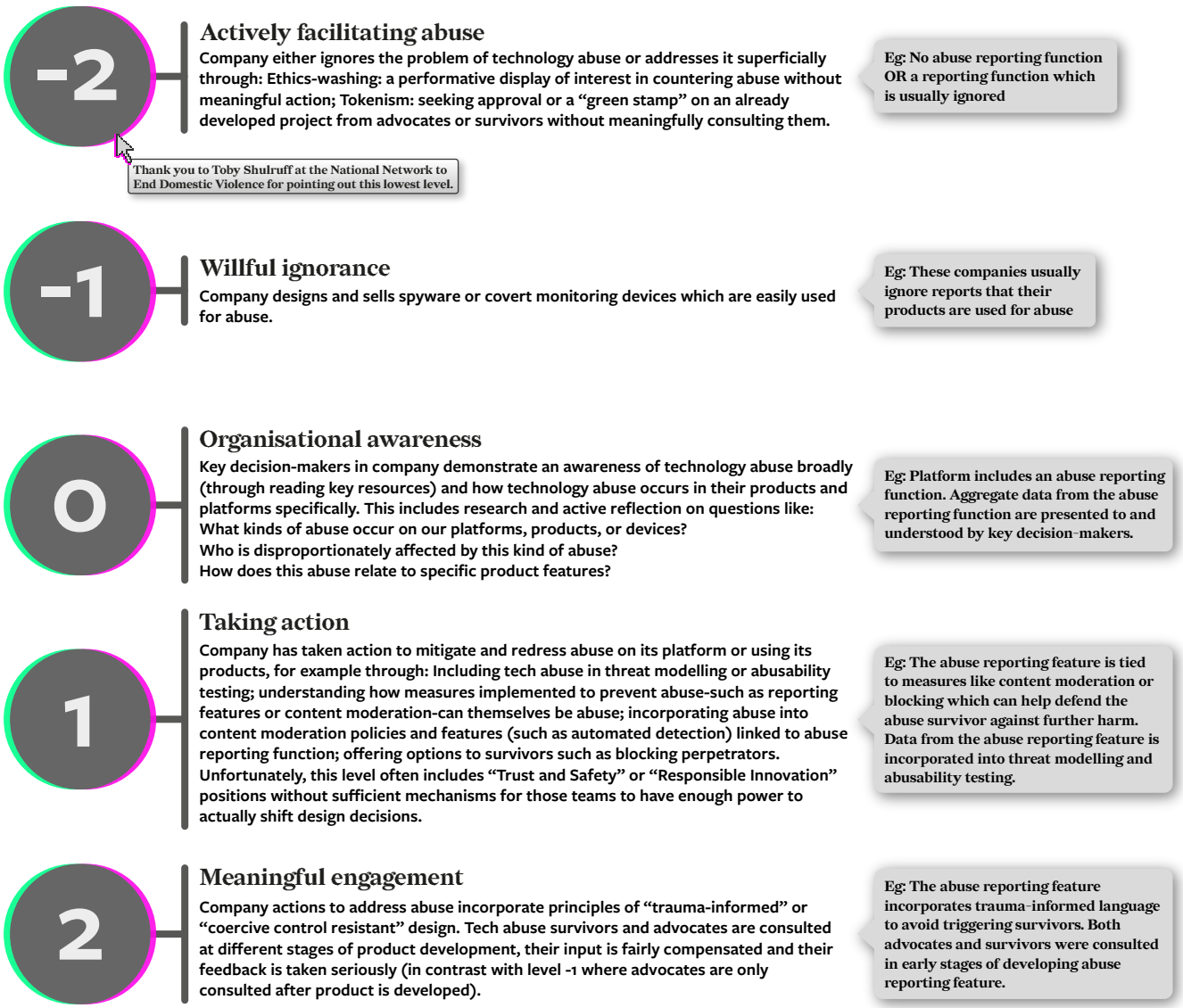
TAKING ABUSABILITY SERIOUSLY

Julia Slupska and Tara Hairston

BUILDING RESPONSIVE SYSTEMS

Although abusability is critical at the design stage (where it is preventative), it is also important to address abuse once it has happened. Both, the advocates who were interviewed for this project and those who attended our workshop, emphasized the importance of building response systems which have a relational dynamic with victim services: responding to problems as they arise and developing support systems for redressing cases of abuse.

The illustration below shows the different steps organisations can take to incorporate notions of abusability into their product design and organisational practices. We highlight that the companies who are perhaps most organizationally mature, might not be the companies who are most mature in relation to their understanding of abusability and their treatment of abuse on their platforms.



Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)

DO'S AND DONT'S FOR JOURNALISTS COVERING TECHNOLOGY-MEDIATED ABUSE

Angelika Strohmayer

- Do understand the differences between threats and harm: Threats are design flaws, weaknesses, or misuses of technology that have the potential to cause harm; whereas harms relate to situations where the harm has already been caused. Distinguishing the two can be important in understanding existing and potential threats, both of which must be handled carefully.

- Do avoid victim blaming: for example when reporting about image based sexual abuse - or revenge porn - the argument should not be that the person of whom the nude photos were taken should not have taken the photos. Instead, the argument should be around consent and that the non-consensual sharing of these images and/or videos is unacceptable.

- Do practice considerate and caring interview techniques. Go at the pace the interviewee is comfortable with, understand when questions become too uncomfortable to continue, and respect the boundaries of your interviewee.

- Do use the power you hold as journalists for good - taking a justice-oriented approach to your reporting.

- Do not sensationalise interpersonal violence, domestic abuse, and other forms of harm mediated with and through technologies. The way stories are written can falsely give the impression that perpetrators of technology-mediated abuse are more technically capable than they are; reinforcing victim survivors' feelings of subjugation

- Do not repeatedly ask interviewees the same question or try to obtain highly personal information if this is at the risk of re-traumatising or in other ways harming the interviewee.

Additional resources for academics wanting to write their own articles: Public Voices Project (https://www.theopedproject.org/) trains researchers on how to write their own stories, articles, and op-eds. The Conversation (https://theconversation.com/) is an initiative that provides space for academics to write news stories about their research.

Additional resources for journalists: Alice Wilder's guide to trauma-informed reporting contains a great deal of information about what to do, what to avoid, and how to best do trauma-informed reporting: https://transom.org/2021/trauma-informed-reporting/ and The Ground Truth project also has a fantastic guide, including 6 tips for interviewing people who have experienced trauma: https://thegroundtruthproject.org/interviewing-people-experienced-trauma/

Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Gina Neff, Lynne Coventry, Tara Hairston, Adam Dodge

EPSRC's Human-Data Interaction Network+ (EP/R045178/1)