



Anticipating the Use of Robots in Domestic Abuse: A Typology of Robot Facilitated Abuse to Support Risk Assessment and Mitigation in Human-Robot Interaction

Katie Winkle
katie.winkle@it.uu.se
Uppsala University
Stockholm, Sweden

Natasha Mulvihill
natasha.mulvihill@bristol.ac.uk
University of Bristol
Bristol, United Kingdom

ABSTRACT

Domestic abuse research demonstrates that perpetrators are agile in finding new ways to coerce and to consolidate their control. They may leverage loved ones or cherished objects, and are increasingly exploiting and subverting what have become everyday ‘smart’ technologies. Robots sit at the intersection of these categories: they bring together multiple digital and assistive functionalities in a physical body, often explicitly designed to take on a social companionship role. We present a typology of robot facilitated abuse based on these unique affordances, designed to support systematic risk assessment, mitigation and design work. Whilst most obviously relevant to those designing robots for in-home deployment or intrafamilial interactions, the ability to coerce can be wielded by those who have any form of social power, such that our typology and associated design reflections may also be salient for the design of robots to be used in the school or workplace, between carers and the vulnerable, elderly and disabled and/or in institutions which facilitate intimate relations of care.

CCS CONCEPTS

• Human-centered computing;

KEYWORDS

trustworthy AI, safety, robot abuse, domestic abuse, feminism

ACM Reference Format:

Katie Winkle and Natasha Mulvihill. 2024. Anticipating the Use of Robots in Domestic Abuse: A Typology of Robot Facilitated Abuse to Support Risk Assessment and Mitigation in Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3610977.3634938>

1 INTRODUCTION

The ‘problem’ of robot abuse has been considered from a number of perspectives within the Human-Robot Interaction (HRI) literature to date. Whilst there has been some discussion on the destruction of robots as creativity or catharsis [44], the majority of works are concerned with the potentially negative impacts of such behaviour.

Regarding the abuse of robots, concerns cited within the literature include the potential for (unchecked) robot-directed abuse to hinder robot function [5], reinforce and/or normalise harmful social norms and behaviours [65, 83], induce distress within human observers of the abuse [25, 61] and/or damage the ‘moral character’ of the perpetrator [13]. Regarding abuse *with* robots, critical stances on the use of robots e.g. in policing and military contexts can be found both in literature [78] and in researcher-led movements like the Campaign to Stop Killer Robots¹ and the No Justice No Robots campaign.²

Seemingly absent from the discussion so far is a detailed reflection on the potential for robots to be (mis)used and abused within the context of domestic abuse. Based on increasing evidence documenting Technology Facilitated Abuse (TFA), including the “subversion of ‘everyday’ digital technological systems to coerce, control and harm” [68, p.4], we must anticipate that robots designed for in-home deployment and/or otherwise explicitly designed to facilitate interactions between family members or partners may also be (mis)used and abused by perpetrators. According to the Istanbul Convention, domestic violence and abuse includes “all acts of physical, sexual, psychological or economic violence that occur within the family or domestic unit or between former or current spouses or partners, whether or not the perpetrator shares or has shared the same residence with the victim” [52]. It can therefore include violence and abuse between “parent and child, siblings, or even roommates” [68, p.13], although these forms of abuse may also be analysed under separate literature on, for example, ‘family violence’ or ‘child-to-parent violence’. Intimate Partner Violence (IPV) is often used interchangeably with ‘domestic violence’ but focuses specifically on abuse between intimate partners, in household, romantic or dating relationships [1]. In this paper, we will use the term ‘domestic abuse’ to refer to violence, abuse and control within intimate partner, familial and/or household settings.

Whilst domestic abuse is both perpetrated and experienced by people of all genders, the ability to coerce is commonly wielded by those who have social power – including e.g. with respect to gender and/or access to particular resources or knowledge, such as wealth and expertise. This can explain why it is disproportionately men who have exerted coercive control in intimate relationships. To give some sense of the scale of the problem: globally, it is estimated that 27% of women in relationships have experienced IPV at least once in their lifetime [53]. In New Zealand, police attend a family harm episode every three minutes [56] with Māori women, nearly 50% of



This work is licensed under a Creative Commons Attribution International 4.0 License.

HRI '24, March 11–14, 2024, Boulder, CO, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0322-5/24/03.
<https://doi.org/10.1145/3610977.3634938>

¹<https://www.stopkillerrobots.org/stop-killer-robots/>

²<https://nojusticenorobots.github.io/>

whom experience partner violence during their lifetime, being at particular risk of harm [55]. In England and Wales, it is estimated that 5.0% of adults (6.9% of women and 3.0% of men) aged 16 years and over experienced domestic abuse in the year ending March 2022; equating to an estimated 2.4 million adults (1.7 million women and 699,000 men) [23]. Specifically concerning TFA, 59% of women and children supported by the UK-based domestic abuse charity Refuge in 2020–2021 experienced abuse involving technology [18, p.52]. In the U.S., a (nationally representative) survey of 2752 adults aged 18–35 found that 70% of all respondents reported experiencing some form of technology facilitated abuse, rising to 81% among sexual and gender minorities, with friends, family and/or acquaintances, together with (ex-)intimate partners, accounting for the largest individual proportion of reported TFA experiences across both groups (43.1% and 61.3% respectively) [71].

When it comes to TFA within domestic settings, the gendered nature of domestic abuse is further compounded by a gender gap in digital skills [75]. Women are less likely to engage with security and privacy technologies (e.g. using VPNs, anti-spyware) than men, which has fuelled calls for work on digital safety protection to be more inclusive of women's needs [17]. Similarly, despite the relative prevalence of TFA, awareness of it remains low. Nearly half of the women surveyed by Refuge were unable to identify a home device they thought could be vulnerable to abuse [18, p.56]. Whilst there has been some shift towards criminalisation of (some forms of) TFA, e.g. in the UK and Australia, it seems unlikely that formal legal sanctions will be sufficient [68, p.7], but rather that there will be a combination of criminal justice and regulatory approaches, with industry being expected to take an active role in these efforts [18, p.56]. Those designing, developing and deploying HRI have an opportunity to get 'ahead of the curve' by considering this now, particularly reflecting on the ways existent smart home/connected devices have been shown to be misappropriated, before e.g. in-home robots become equally as pervasive. Risk assessment and mitigation seems an appropriate method for conducting such work in a systematic and evidence-informed way [6, 15], yet there exist few conceptual tools to help support such a process. This simultaneously reflects and exacerbates the fact that general awareness surrounding domestic abuse remains low, in part due to a chronic history of being overlooked [37]. In this article we bring together insights across HRI, domestic and intimate partner violence and existent mechanisms of TFA, to create a typology of robot facilitated abuse. We situate our work as supporting both Trustworthy AI [15] and Feminist HRI [82].

1.1 Supporting Trustworthy Robotics/AI

Whilst a number of robot ethics guidelines and frameworks might be used to motivate this consideration of robot facilitated abuse, we specifically posit our work as an effort to support those designing, developing and deploying HRI in practically realising the EU Ethics Guidelines for Trustworthy AI [15]. Firstly, under the principle of 'prevention of harm', the EU Ethics Guidelines for Trustworthy AI call for developers to ensure their systems are not open to 'malicious use', with particular attention being paid to situations where they might "cause or exacerbate adverse impacts due to asymmetries of power or information" [15, p.12]. The notion of a 'malicious actor'

appears again under the requirement for technical robustness and safety: "*possible unintended applications of the system... and potential abuse of the system by malicious actors should be taken into account, and steps should be taken to prevent and mitigate these*" [15, p.16]. Considering the ways in which robots might be intentionally used to enact domestic abuse represents exactly such an exercise, motivated by the ways in which we know other smart and digitally connected devices have been used.

The EU guidelines refer to protecting "*human dignity, as well as mental and physical integrity*" and note that "*harms can be individual or collective, and can include intangible harm to social, cultural and political environments*" [15, p.12]. However, the concept of 'harm' still remains relatively abstract. We suggest a zemiological approach/framing aligns well to the guidelines' broad conceptualisation of harm, whilst also helping us to identify explicitly the types of harm to be considered in robot deployment. In contrast to criminology, which tends to be concerned with state-led definitions of crime and criminalisation, a zemiological approach aims for social justice. It allows for identification of - and potential to address - harms that would otherwise go unseen or under-recognised: "*the primary motivations of zemiology are therefore to provide conceptual tools to focus on harms which are endemic, but that often pass unseen and indeed unregulated, and which are themselves often based in power imbalances and intersectional oppressions*" [7, p.33]. In short, just as the guidelines indicate compliance with legal requirements is necessary but not sufficient to deliver trustworthy AI, zemiology asserts that (significant) harm can be enacted even where no (illegal) criminal activity occurs. This allows for critical discussion of international and mass harm (e.g. climate change) as well as individual harms which are facilitated by practices and omissions at the macro-level. Such harms may not be amenable to definition in criminal law or be effectively deterred or impacted by criminal sanction, because their causes are systemic, as has been argued for violence against women [7, p.27] and other forms of technology-facilitated harms [26].

A typology of harms pertinent to the zemiological approach might include physical harms, emotional and psychological harms, financial and economic harms, cultural harms, harms of recognition and autonomy harms [7, p.66]. Such an approach explicitly supports better harm anticipation and reduction, analogous to the notion of ethical risk assessment and mitigation [6, 15]. We limit our considerations in this paper to physical, sexual and psychological harms that may arise from perpetrator-robot-victim interactions, as we find these to be particularly relevant to HRI research/design in the context of robot facilitated domestic abuse (compared e.g. to other forms of TFA). However, we encourage HRI researchers interested in trustworthy, safe and ethical HRI to consider making use of the breadth afforded by the zemiological approach when considering potential harms associated with robot deployment.

1.2 Supporting Feminist HRI

A feminist approach to HRI calls foremost for those designing and developing HRI to consider how its deployment might impact/be impacted by the power dynamics at play between different actors. These include robot users and those around them, as well as the institutions and social, cultural, legal structures in which they are

embedded [82]. Particular attention should be paid to ‘low power users’ who may be most at risk of negative consequences resulting from robot deployment. Indeed, an increasingly activist design stance may look to centre the needs and experiences of these users, challenging the status quo of robot deployment and legitimising work which takes seriously the (embodied) experiences of those traditionally under considered. As previously alluded to, domestic abuse is often underpinned by power differentials between perpetrators and victim/survivors. Our typology supports work both ‘examining’ and ‘challenging’ such dynamics. It supports risk assessment and reflection on the former, and design of mitigating strategies - perhaps even new design opportunities - on the latter. However we want to stress the need for caution in any such work and encourage the pursuit of a reflexive design process as put forward by the Feminist HRI approach (see more under Section 4).

2 TECHNOLOGY FACILITATED ABUSE

Technology Facilitated Abuse (TFA) is a broad umbrella term for abuse enacted and/or exacerbated via technological means. Examples include, e.g. cyberstalking, online harassment and abuse related to gender or sexuality, cyberbullying, and financial fraud. As a form of interpersonal violence, TFA uniquely “*extends the reach of potential abusers, can be perpetrated without direct contact with victims, and may offer a degree of anonymity to perpetrators*” [71]. In particular, we focus here on the concept of coercive control, a commonly observed behaviour within domestic abuse [48, 66, 68, 86]. Much of the literature concerning coercive control relates to intimate partner violence, that is enacted by current and former spouses or dating partners. However the ability to coerce can be wielded by those who have any form of social power – whether in terms of gender, ethnicity or age – and/or access to particular resources or knowledge – such as wealth, professional qualifications or expertise, secure citizenship and residency, gatekeeper status, or workplace seniority. Accordingly, the behaviours, typology and design reflections that we present later in the paper and developed in relation to the domestic and intimate partner context, may also be salient to application contexts in which there exist significant power differentials between different robot users. These include in the school or workplace; between carers and the vulnerable, elderly and disabled; and/or in institutions which facilitate intimate relations of care, mutual support or supervision.

Within an intimate partner context, Stark [66] identified “*a pattern of behaviours by the perpetrator, which cumulatively undermines the personhood and restricts the freedom of their victim*” (see also [48]). These behaviours may include physical, sexual, emotional, psychological and/or financial abuse and threats as well as tracking, harassing and isolating victims to bring them within the perpetrator’s exclusive sphere of influence and control. Increasingly, research is recognising how perpetrators are agile in finding new ways to coerce and to consolidate their control. As ‘smart’ and connected devices have become increasingly pervasive within the domestic environment, evidence suggests that TFA has become “*integral to coercive control*” [88]. A review of UK domestic abuse court cases found that, within domestic abuse, TFA was centred on control and manipulation, noting that these findings echo other studies in suggesting that the purpose of such abuse is “*rarely to*

access data (except where there was the added element of control) but rather to coerce and monitor victims/survivors” [68, p.37]. In describing ‘digital coercive control’ Woodlock et al. [86, p.369] note how technology enables perpetrators to exercise new forms of “*intimidation, isolation, shaming, micromanagement of daily activities, and surveillance*”. The embeddedness of technology in everyday life also means that perpetrators can project omnipresence, “*with victims feeling they can never really escape the perpetrator’s abuse*” [86, p.369] (see also [85]).

Sexual coercion, violence and abuse is an under reported yet common element of domestic abuse. Anastasia Powell and Nicola Henry began writing on what they termed ‘technology facilitated sexual violence’ (or TFSV) in the early 2010s. They noted that, “*...smartphone applications, dating websites, social media platforms and intimate digital images are all being used to facilitate sexual assaults as well as to shame, humiliate and blackmail victim-survivors*” [57, p.79]. Powell and Henry were particularly keen to foreground the embodied nature of the harm experienced through TFA, and in this way reject the traditional distinctions between ‘body and mind’ and ‘online and offline’ [29]. Taking seriously the material harm of intangible practices is increasingly reflected in laws, for example, around ‘revenge porn’ or ‘online harassment’, and in some ways echoes the shift in recognising forms of intimate partner violence beyond physical violence.

Powell and Henry have also recognised how such practices “*represent extensions of sexual violence, often amplifying the impacts on victims, rather than representing necessarily a new or unfamiliar harm*” [28]. This fits with the ‘computer-assisted crimes’ category in David Wall’s 2001 typology [74] which distinguishes between existing crimes facilitated in new ways versus entirely new computer-dependent crimes. However, there is also a political point to identifying TFSV as part of the continuum of sexual violence [39] because it recognises that, within new technologies, we continue to remake and re-affirm long-standing social hierarchies, including gendered hierarchies. For that reason, the potential for harm should be expected and mitigated against (we note this aligns to existent HRI work concerning, e.g. robot-directed abuse normalising problematic behaviour [81, 83]).

Drawing on evidence across academia, government and public services, a recent UK parliamentary report documents the ways in which ‘connected devices’ are currently being exploited for/used to exacerbate behaviours and patterns of domestic abuse [18]. Specifically mentioned are smart home security systems; smartphones and tablets; wearables; and other smart home devices such as connected toys, baby monitors, cameras and speakers. Key report insights include [18, p.53-54] (authors’ summarisation):

(1) Perpetrators are often the household members who purchase, set up and/or manage devices, often controlling Wi-Fi settings within the home and/or forcing victim/survivors to divulge passwords to their accounts.

(2) Perpetrators can use connected devices to overtly coerce and control victims/survivors, even when they are not physically present. This may include persuading and/or misleading victim/survivors that devices can perform certain activities (e.g. record video or audio) that they in fact cannot.

(3) Perpetrators can use connected devices to covertly monitor victim/survivors and collect, e.g. recordings or images to be used

in blackmail. Victim/survivors may underestimate the capabilities of devices they interact with. Remote access features may enable perpetrators to continue abuse after victim/survivors have fled.

(4) Perpetrators can use connected devices for retaliation against victim/survivors. This can include manipulating devices and/or the victim/survivor's material environment in order to 'gaslight' and cause fear, confusion and distress. Geolocation features may compromise the victim/survivor's physical safety by disclosing their location and route history.

These points provide a useful start for thinking about robot facilitated abuse, but we suggest there are gaps and specificities in relation to the use and abuse of robots, which deserve special attention.

2.1 Robot Facilitated Abuse as a Special Case of Technology Facilitated Abuse

We identify three key reasons why robots are likely to afford new and/or exacerbated forms of TFA and hence require dedicated consideration from a product design, risk assessment and mitigation perspective. First, the bringing together of multiple digital/assistive functionalities into one single device is likely to increase ease of (perpetrator) access to those functions, i.e. the robot is likely to afford multiple forms of TFA we currently see facilitated by smart home and connected devices. Second, we know that perpetrators can leverage, for example, children, companion animals, medicines, religious belief, cherished objects or activities, to intimidate and punish their victims [48, 67, 72, 73]. Compared to other technological devices, robots deployed within the home may represent particularly 'cherished objects' and/or, depending on the level of anthropomorphism³, potential social companions. This can make them a target for perpetrators. Third, teleoperable robots - like the Double⁴ - also facilitate new and unique forms of embodied, robot-as-physical-avatar human to human interaction. Finally, all of these robot-specific features intersect and compound. For example, the risks associated with the robot as cherished object or social companion may be exacerbated by the robot's potential to support surveillance and perpetrator 'omnipresence', as shown in existing forms of TFA.

2.1.1 The (Anthropomorphic) Physicality of Robot Bodies. Most TFA is 'device agnostic': this means that it can be enacted without the need for any (one) specialist device to be installed within the victim's environment (e.g. cyberbullying or sexual harassment via social media or direct messaging). However, where there is one or more physical and identifiable embodiments associated (by the victim/survivor) with TFA (consider, e.g. home surveillance systems with visible cameras, smart speakers or home assistants), these can serve as a powerful, physical reminder of a perpetrator's (potential) omnipresence [85]. The physical presence of such devices can be enough to impact on victim/survivor behaviour, even in the absence of functioning digital capabilities (c.f. Mark Coeckelbergh's reflection on the ability of a non-functioning speed camera to regulate driver behaviour [14, p.13]). Further, the physicality of the

robot may further be explicitly exploited by perpetrators wanting to unnerve victims/survivors. It may be placed in specific areas of the home in order to generate a sense of surveillance and/or perpetrator omnipresence, or physically positioned in ways designed to generate discomfort (perhaps leveraging something similar to the uncanny valley effect [47]).

Secondly, the robot body can itself be a target of abuse. This is complicated by the ways in which robot bodies (and 'identities' [46, 80]) are situated in material-discursive practices which "*reflect and influence structures of power*" (see discussion in [82]). Particular robot designs may reflect particular norms or ideals, and interactions with the robot (whether positive, negative, violent etc.) may be seen as a proxy engagement with these norms or ideals. This underpins a key argument for why, e.g. the "rape" of humanoid sex robots may be problematic [60, 65], and has motivated work exploring how responses to robot-directed abuse might best avoid propagating harmful stereotypes, such as women acquiescing to poor treatment [81, 83]. Generally, previous work has discussed these issues in the context of influencing societal norms 'at large'. Our focus in this paper is empirically addressing harm to individuals, albeit we recognise the interconnection of individual practices and socio-cultural norms. We therefore build on previous works that have considered distress induced in those who observe physical abuse of robots [25, 62]. Particularly relevant here is Hideki Garcia Goo and colleagues' recent intersectional study on distress induced by observations of robot-targeted abuse [25]. The authors found a positive correlation in the amount of distress induced where participants were female or a marginalised gender identity, and had previous experience of victimisation. They found negative correlation (i.e. minimal or no distress) for participants who agreed with sexist and/or anti-egalitarian viewpoints. Their insights point directly to the sort of risks we explore in this paper, as well as the need for mitigation by roboticists: "*we might anticipate a scenario in which a man abuses a female-presenting robot, with no negative (emotional) consequences to himself, whilst causing harm to witnesses of the interaction*" [25, p.2445].

We would argue that, whilst anthropomorphism may increase certain risks associated with robot-directed abuse, the potential for harm is not *predicated* on it. That is to say, the risks of robot-directed abuse are not limited to robots of particularly anthropomorphic design. Consider the targeted, intentional destroying of someone's smartphone. Smartphones afford communication with both social and emergency contacts; they often store photos, videos and messages of sentimental value, and might also be used to capture evidence of abusive behaviour as a step towards seeking legal help [18]. In destroying the phone, an abuser demonstrates (both literally and symbolically) their capacity to isolate/control its owner and a potential willingness to engage in violent behaviour. In summary, whilst anthropomorphism is not a necessary ingredient for leveraging the robot's physicality in causing harm, it can be an additional, even exacerbating way, to do so, unlikely to be covered by existent TFA literature.

2.1.2 Robots as Social Companions. Robots for in-home deployment are often explicitly designed to take on social companion roles as part of, or in addition to, their primary function. They may

³We refer here specifically to the way in which user(s) may anthropomorphise the robot and consider it e.g. as a 'friend' rather than any specific physical and/or behavioural design intent.

⁴<https://www.doublerobotics.com/>

present a carefully crafted anthropomorphic or zoomorphic persona, designed to maximise user-robot engagement (e.g. [12, 20, 34]). We do not wish to reinforce an idea that in-home robots ought to be anthropomorphic by default and we would point to existing work problematising and exploring the ethics of anthropomorphic design [64, 79]. However, we would note that real-world deployments of robots indicates the extent to which users might anthropomorphise and/or become attached, even to relatively non-anthropomorphic/zoomorphic robots [8, 9, 24]. Anthropomorphism seemingly plays a role in user-robot emotional connection [19]. It is this potential for emotional connection (whether maximally curated or not) that may increase the potential for harm arising from abuse against robots. Take the harrowing case of UK-based care workers tormenting dementia patients by ‘harming’ their comfort dolls; simple, soft toy baby dolls that patients care for as if they were real babies:

“In one photo, titled ‘Tug of War’ [carer 1] pulls a doll off an elderly woman who holds on with her hands to stop it being snatched away... on another occasion [carer 2] filmed [carer 1] throwing a doll to the floor and shouting “Die baby. Die.” [49]

It seems reasonable to expect that the patients’ distress might have been increased, should their dolls have been equipped with social robot capabilities like anthropomorphic movement and emotive communication (and, as a consequence, the potential to go quiet, stop moving and, perhaps even more convincingly ‘die’). Admittedly, the nature of the patients’ illness may have meant they were more vulnerable to confusion as to the doll being truly alive. But we would again draw attention to the very real emotional connection that users develop towards robots which they know are ‘just’ machines [30]; to the distress that can be induced by observing robot directed abuse; and to the symbolism of destroying something/some “one” cherished by the victim/observer.

We may also anticipate ways in which perpetrators might maliciously interfere with robot social companions in order to have the robot - from the victim/survivor’s point of view - deliver abuse. Direct teleoperation, including e.g. the ability to have the robot vocalise particular speech, represents one way such (mis)appropriation might occur. We would also point to end-user programming, personalisation and machine learning approaches (e.g. [63]) as presenting the same/similar risks. Consider, for example, a localised version of Microsoft’s Tay [84], purposefully ‘taught to be bad’. The victim may or may not understand that such robot behaviour actually originates from the perpetrator (especially given previously mentioned digital knowledge disparities and the associated potential for intentional gaslighting). This is then exacerbated by potential user-robot attachment and/or reliance on the robot for its assistive functionalities. We might hypothesise that abuse from a robot, with which an individual has developed an emotional attachment, could be experienced as a form of ‘social betrayal’. We note that ‘betrayal trauma’ [21] can provoke severe distress, reluctance to recognise the abuse, and difficulty in forming future attachments.

2.1.3 Robot as (Physically Embodied) Avatar. We have already discussed the physical embodiment of robots exacerbating the sense of perpetrator omnipresence, in a similar way to surveillance cameras or smart speakers. However, the potential for real-time robot body teleoperation brings a new dimension to this, which is

unique to (mobile) robots over other devices that might be installed within the home. Documented incidents of sexual harassment in virtual reality [3, 76], indicate we must also consider the risks to victim/survivors who make use of teleoperated robots within domestic, familial and/or intimate partner settings. Extant research on violence - including sexual violence - in virtual and augmented reality spaces, has tended to focus on the moral and ontological status of violence [89] and the appropriate criminal justice [69] or ethical design [22] responses. Commonly, authors seek to compare and contrast the harms of ‘in-person’ versus ‘virtual violence’ and to explore to what extent both are connected, including causally. We propose that robots-as-avatars can be used to coerce and abuse both as an independent tactic and in a way that consolidates in-person abuse.

3 A TYPOLOGY OF ROBOT FACILITATED ABUSE

Bringing together insights from literature spanning domestic abuse and intimate partner violence, existent mechanisms of technology facilitated abuse and HRI, we present a typology of robot facilitated abuse covering three perpetrator-robot-victim configurations in Table 1. These configurations were derived iteratively, drawing on our respective research expertise (KW: HRI, NM: Criminology), in an approach akin to critical interpretive synthesis [2]. We reviewed (a) HRI literature concerning robots to be used within the home (e.g. [10, 20, 24, 35, 40, 42, 45]) and/or designed to facilitate interactions between family members/partners (e.g. [4, 50, 70, 87]); (b) literature on TFA [18, 29, 57, 68, 71, 86, 88]; and (c) existing and emerging work on domestic abuse and coercive control [48, 67, 72, 73]. We started from a conceptualisation of in-home robots as assistive, digital ‘smart-home-like’ devices, with potential to become cherished objects and/or social companions, and hence at risk of being exploited by perpetrators. The typology is necessarily an exercise in horizon-scanning, but it is rooted in reflexive and critical analysis of existing knowledge.

Robots discussed in HRI literature range from those that are (identifiably) human teleoperated to those which are designed to present a particular (seemingly autonomous/agentive) social ‘character’. Therefore, we identified a need to distinguish between robots as avatars versus robots as social agents. In the typology, our *Robot as Victim* and *Robot as Perpetrator* configurations relate primarily to the latter category, but notably refer to the way the robot is presented, perceived and/or conceptualised, rather than the specific mode of operation/level of autonomy. We group *Robot as Avatar* together with *Robot as Tool*, since both concern the leveraging of robot functionalities as an instrument for enacting abuse.

In Table 1 we suggest examples of physical, sexual and psychological abuse that we might anticipate across these three configurations within the context of domestic abuse. Both the configurations and example behaviours we outline should not be considered mutually exclusive nor exhaustive. Rather, they are an evidence-informed starting point designed to support HRI practitioners in systematically identifying risks, mitigation strategies and design opportunities pertaining to robot facilitated abuse in domestic, familial and/or intimate partner contexts. It is crucially important to note that an

interaction might be construed differently by perpetrators and victim/survivors, potentially in ways that the design team might have thought unlikely. In Table 1, we speak primarily from a perpetrator perspective, suggesting ways in which perpetrators might look to leverage robots in enacting abuse. We do this first, to underscore robot facilitated abuse as abuse knowingly and willfully enacted by perpetrators, for which they ought to be held accountable. Second, we think this is a sensible starting point for thinking about design mitigations, aiming in the first instance to try and *prevent* such (mis)use and abuse of robots [25]. However, we should also acknowledge that there will be cases where prevention is impossible/insufficient, so we must also work to *minimise* harm where such abuse does occur. Consideration of the victim/survivor perspective should be centred here. For example, the difference between the *Robot as Perpetrator* and the *Robot as Avatar* configurations is that, in the former, the perpetrator has the robot *character* (appear to) deliver the abuse. If the victim/survivor knows and understands the robot's behaviour to be coming from - or programmed by - the perpetrator, they may conceptualise the robot more as an extension or avatar of the perpetrator, and hence experience per the *Robot as Avatar* configuration. Therefore, while the *Robot as Victim* or *Perpetrator* configurations might seem particularly relevant for social and/or highly anthropomorphised robots, and the *Robot as Avatar* may seem particularly relevant to robots designed primarily for teleoperation, the reality is that these configurations are fluid and that they depend on perpetrator behaviour and victim perception. We try to exemplify this in Section 3.1.

3.1 Example Typology Application

It is important to stress that our typology is, to some extent, an exercise in anticipating futures, albeit rooted in peer-reviewed evidence of what we know/see already in the domestic abuse and TFA spaces. As with most risk assessment and mitigation work, reflecting on the risks identified here is not guaranteed to identify all possible ways a particular robot/application might be (mis)used and abused, but it does offer a tangible starting point. To give some indication of how, we use it here to identify, compare and contrast risks posed by two different in-home robots: Roomba and Jibo. The Roomba robot is an autonomous robot vacuum cleaner which can be partially controlled (e.g. directed to clean a particular room or space at a particular time/on demand) via the iRobot Home App.⁵ Jibo is a socially assistive robot which - according to promotional materials from its time as a commercial product⁶ - can capture photos and videos, relay messages from friends/family and support remote teleoperation for video calls. Recent work posits using Jibo to delivering in-home psychological wellbeing interventions [34, 36] and it is this application we reflect on here.

Looking at our typology, we might predict that perpetrators might physically damage Roomba per the *Robot as Victim* or *Robot as Tool* configurations. The former might be done to induce distress or fear in the victim/survivor (noting that there is evidence of users developing emotional attachment to service robots [8, 9, 24] but again that observations of violence are distressing regardless of specific user-robot attachment [25]). The latter might be done

to exert control over the victim/survivor's behaviour, i.e. forcing them to do the cleaning the robot was intended to do. In cases where perpetrators have access to the app, they might achieve similar via interfering with the robot's operation (e.g. cancelling scheduled cleans). They might also use app access to schedule cleans at inappropriate times or act in unexpected ways, potentially enacting psychological abuse per the *Robot as Tool* configuration.

In comparison, we might suggest that Jibo is more at risk of being leveraged per the *Robot as Avatar* configurations given its teleoperation and message delivery functionalities. We would also suggest the use of Jibo for psychological wellbeing specifically generates increased and intersecting risks across the *Robot as Victim*, *Robot as Perpetrator* and *Robot as Tool* configurations. Engagement with the robot-delivered intervention likely relies on some level of user-robot attachment (per the therapeutic alliance/therapeutic relationship [32, 43]) such that abuse, towards or from the robot, could be more distressing than in the Roomba case. Similarly, interference with Jibo's primary therapy-delivery function (whether, e.g. through physical damage per the *Robot as Victim* configuration or simply through interference, e.g. with data saves, memory wipes etc.) has potential to cause significant harm/distress to someone engaged in the intervention.

We note with these two examples how both more and less social/anthropomorphic robots may be leveraged across all of our typology configurations, and that risks may mutually intersect and compound.

3.2 A Cautionary Note on Designing for Risk Mitigation

Designing for safety/security of robots and user protection is a basic design obligation and there are existing efforts to identify technology design principles for risk minimisation (see e.g. those by IBM [51]). However, we want to note that design for risk mitigation is generally analogous to the concept of "situational crime prevention" in criminology, often based on rational, technical beliefs, which may have unintended consequences (displacement, new harms, people behaving in less 'rational' ways than expected resulting in minimal actual deterrence) [31]. In addition, this approach tends to individualise the problem (and responsibilise the individual), when issues such as gender-based violence have structural and systemic causes. So whilst a principled design approach is important and worthwhile, we must recognise that in isolation this will not be sufficient to tackle the issue of robot facilitated abuse. We encourage designers to look through e.g. the IBM guidelines, but note that more work is needed to expand on and translate these towards robots and HRI specifically.

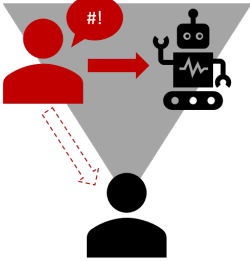
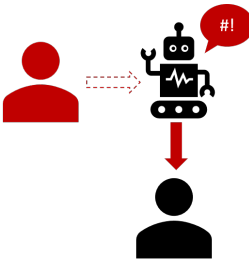
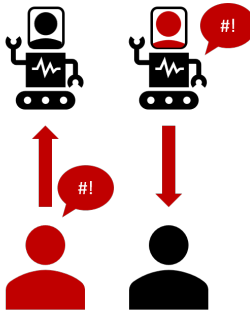
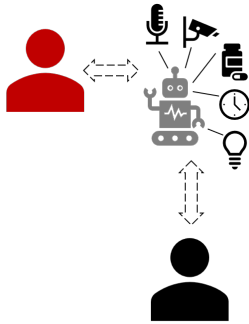
4 BEYOND MINIMISING HARM: SUPPORTING VICTIM/SURVIVORS AND TACKLING ABUSE

Readers of this article may wonder whether there could be a more active role for robots and HRI in tackling domestic abuse. We would urge researchers to think carefully about if, how and why they might consider positing an HRI-based intervention for domestic abuse, noting critique of digital technology interventions for sexual and gender-based violence prevention and response [16]. However, we also see positive developments in other technological fields, and

⁵https://www.irobot.com/en_US/irobot-home-app.html

⁶<https://www.youtube.com/watch?v=H0h20jRA5M0>

Table 1: A typology of robot facilitated abuse presenting three perpetrator-robot-victim configurations. Examples of physical, sexual and psychological ("psych.") robot facilitated abuse we might anticipate within settings of domestic abuse, are provided.

	Robot as Victim (or Cherished Object)	Robot as Perpetrator	Robot as Instrument as Avatar	Robot as Instrument as Tool
	 <p>Perpetrator abuses robot as a way to harm the (human) victim/survivor (explored in previous HRI works [25, 61]). Victim/survivor may or may not be observing in real-time.</p>	 <p>Perpetrator programmes or otherwise controls the robot to behave abusively to the victim (i.e. abuse appears to come from the robot, c.f. Microsoft Tay [84]). Victim may or may not know/understand that the perpetrator is directly responsible for the abuse.</p>	 <p>Perpetrator asserts self and enacts abuse via the robot body, something we expect given demonstrations of similar via VR avatars [3, 76]. This includes being abusive to the victim/survivor-robot body (left) or using the robot body to deliver abuse (right), the latter differing from <i>robot as perpetrator</i> because here the (human) perpetrator is openly the aggressor.</p>	 <p>Perpetrator utilises digital and/or assistive functionalities of the robot to enact abuse in ways they might utilise any other technology c.f. the existent literature on technology facilitated abuse (TFA) across e.g. smart home security systems, smartphones/tablets, wearables, baby monitors, and smart speakers [18].</p>
Physical Abuse	e.g. punching, kicking, throwing the robot	e.g. having the robot strike the victim, or otherwise cause physical harm (e.g. tripping) through its (mis)behaviour	e.g. punching, kicking, throwing the victim-teleoperated robot, or using the teleoperated robot body to strike the victim	e.g. maliciously interfering with operation or physical structure of the robot in order to cause harm through (mis)function (consider e.g. healthcare)
Sexual Abuse	e.g. seeking physical sexual gratification using the robot (without consent of victim-observer), directing sexualised violence towards the robot	e.g. having the robot sexually harass the victim by making inappropriate comments, engage in unwanted touching and/or behave in other sexualised ways	e.g. sexually harassing the victim-teleoperated robot body (e.g. via unwanted touch, indecent exposure), or using the teleoperated robot body to deliver unwanted sexual advances, harassment or violence	e.g. using the robot's audio-visual and/or internet capabilities to capture, upload, access, download or broadcast sexual content with the intent of harming the victim
Psych. Abuse	e.g. verbally abusing the robot (slurs, taunts), or denigrating the victim-robot relationship	e.g. having the robot verbally abuse the victim (engaging in gaslighting, bullying etc.) or work to undermine the victim's social relationships and reduce their contact with friends/family	e.g. delivering verbal abuse (engaging in gaslighting, bullying etc.) via the teleoperated robot, and/or using the robot body to create a sense of perpetrator omnipresence	e.g. using (mis)using or maliciously interfering with robot functionality in order to interfere with the victim's environment in ways designed to induce fear, confusion and/or distress

recognise that such work could represent a compelling exercise in ‘challenging power’ per Feminist HRI [82]. Here, we outline some possible future directions based, in part, on efforts in related fields.

(1) *In what ways could robot-collected/HRI-generated data signal domestic abuse? In what ways could this prompt an intervention or support victim/survivor evidence gathering?* For example, machine learning for healthcare researcher Irene Chen has explored the possibility of predicting intimate partner violence using radiology reports [11]. Importantly, Chen’s work is conducted collaboratively with medical experts, and the aim is not necessarily to create an autonomous classification system but rather empower clinicians to spot potential cases. We would draw particular attention to the way Chen and colleagues have considered the way this AI-based system would fit into (and potentially address weaknesses of) existing structures of care: “Patients are reluctant to come forward...and also clinicians and healthcare practitioners are often not on the lookout for these things. They’re not trained in the right way. They might be busy, they might be resource constrained. They might only have a few minutes with each patient at a time” [54]. Perhaps more relevant for in-home HRI are efforts to identify domestic abuse through analysis of ambient sensor data (see [41] for a review). Notably, in our typology, we have focused only on robot-directed abuse through the lens of (immediate) harm to victim/survivors. Even in cases where there is no immediate victim/observer, we might also discuss whether sustained robot-directed abuse might represent one ‘red-flag’ risk factor indicating the potential for domestic abuse. Research tells us that evidence of ‘jealous surveillance’ [59], monitoring, threats and isolation of the victim, are core tactics of coercive control and domestic abuse. We know also that perpetrators will use whatever tools are at their disposal [48]. Therefore, it is reasonable to infer that the abuse of robots in the presence of an intimate other, as defined in the typology, could be a perpetrator red-flag. Independent abuse of robots, without the presence of an intimate other, may also be a warning sign, though further research is needed. Overall, we must be careful not to lose sight of victim/survivor agency with respect to when (not) and how (not) to report domestic abuse and/or request support. Mobile applications are currently available for victim/survivors to support, for example, documenting and/or recording abuse, anonymously (or not) contacting law enforcement, and finding professional support. Such efforts must be careful to ensure any data recorded is usable by authorities (e.g. there may be a need to ensure data is time-stamped) and should avoid propagating the idea that “surveillance is a normal approach to protecting women” [58]. Further work on this would require input from victim/survivors, related advocacy groups/service providers and law enforcement, and may be locally constrained due to differences in legal definitions and classifications of domestic abuse as a crime.

(2) *Could HRI aid in tackling some potential “root causes” of domestic abuse, e.g. by supporting (would-be) perpetrators?* For example, ‘innerBoy’ is a mobile application designed to support men in healing from the impacts of intergenerational trauma and abuse created by Taimalelagi Mataio (Matt) Brown, survivor and founder of “She Is Not Your Rehab”.⁷ innerBoy is designed for “men who are often left out of the conversation in family violence prevention with limited

non-mandated therapeutic options available to them”, essentially providing a tool for ‘self-therapy’. The idea of using socially assistive robots to support such therapy seems like an obvious possibility given existing demonstrations of such robots in supporting psychological well being exercises [34]. Using robots to facilitate discussions between men, supporting delivery of something like the “#guytalk” initiative⁸ (conceived by Swedish equality consultants MakeEqual, #guytalk facilitates conversations about masculinity) would seem to extend existing work on the use of social robots as mediators of difficult conversations (e.g. [33, 38, 50, 70]).

(3) *Beyond trying to tackle/minimise domestic abuse directly, how else could the HRI community support work to address domestic abuse?* Coercive control prevention efforts generally encompass improving knowledge and awareness, reducing shame and fear and improving confidence in reporting. Raising knowledge and awareness among the robot design community and among manufacturers is an important step in working to minimise (mis)use of robots of the kind we have seen with other smart/connected devices. The idea of robot facilitated abuse may currently seem outlandish to many, so early signalling by the HRI community and linking to existing forms of coercive control that people recognise (e.g. phone trackers), may help to embed and legitimise this as an issue to be taken seriously. On the consumer/user side, given the documented gender divide in digital skills [75], continuing to work on better engaging girls in computing education and robotics [27] might help to reduce knowledge disparities that might be leveraged in robot facilitated abuse. This should be part of broader efforts to educate young people on the ethical risks associated with new technologies [77].

Overall, our greatest source of expertise here is going to be victim/survivors themselves, so we should be making sure to involve them (safely, and properly compensated) within our design and development processes. Sometimes, what may appear as a good and evidence-backed idea, could turn out to be unsafe or unworkable. There is no substitution for the perspectives and lived experiences of victim/survivors in identifying and addressing such shortcomings.

5 CONCLUSION

We must anticipate that robots developed for use in domestic, familial and/or intimate partner settings may be (mis)used and abused by perpetrators of domestic abuse. We have provided a typology of robot facilitated abuse, identifying possibilities uniquely afforded by robots over other technologies currently exploited for technology facilitated abuse. We hope this work serves to legitimise further study of robot related and domestic abuse within HRI, and provides a practical tool for systematic risk assessment, mitigation and design work in this context.

ACKNOWLEDGMENTS

This work was partly funded by the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

⁷<https://www.sheisnotyourrehab.com>

⁸http://www.killmiddag.se/index_eng.html

REFERENCES

- [1] Women Against Abuse. [n.d.]. The Language We Use. <https://www.womenagainstabuse.org/education-resources/the-language-we-use>. Accessed: September 27, 2023.
- [2] Elaine Barnett-Page and James Thomas. 2009. Methods for the synthesis of qualitative research: a critical review. *BMC medical research methodology* 9, 1 (2009), 1–11.
- [3] Tanya Basu. 2021. The metaverse has a groping problem already. *MIT Technology Review* (2021).
- [4] Andriana Boudouraki, Stuart Reeves, Joel E Fischer, and Sean Rintel. 2022. Mediated visits: Longitudinal domestic dwelling with mobile robotic telepresence. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [5] Dražen Brščić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from Children's Abuse of Social Robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. Association for Computing Machinery, New York, NY, USA, 59–66. <https://doi.org/10.1145/2696454.2696468>
- [6] BSL. 2016. BS 8611:2016 Robots and robotic devices: guide to the ethical design and application of robots and robotic systems.
- [7] Victoria Canning and Steve Tombs. 2021. *From Social Harm to Zemiology: A Critical Introduction*. Routledge, London. <https://doi.org/10.4324/9780429430497>
- [8] Massimiliano L Cappuccio, Jai C Galliot, and Eduardo B Sandoval. 2021. Saving private robot: Risks and advantages of anthropomorphism in agent-soldier teams. *International Journal of Social Robotics* (2021), 1–14.
- [9] Julie Carpenter. 2016. *Culture and human-robot interaction in militarized spaces: A war story*. Routledge.
- [10] Julie Carpenter, Joan M. Davis, Norah Erwin-Stewart, Tiffany R. Lee, John D. Bransford, and Nancy Vye. 2009. Gender Representation and Humanoid Robots Designed for Domestic Use. *International Journal of Social Robotics* 1, 3 (March 2009), 261. <https://doi.org/10.1007/s12369-009-0016-4>
- [11] Irene Y Chen, Emily Alsentzer, Hyesun Park, Richard Thomas, Babina Gosangi, Rahul Gujrathi, and Bharti Khurana. 2020. Intimate partner violence and injury prediction from radiology reports. In *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*. World Scientific, 55–66.
- [12] Chatchai Chirapornchai, Paul Bremner, and Joseph E Daly. 2021. Helper's High with a Robot Pet. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 229–233.
- [13] Mark Coeckelbergh. 2021. How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance. *International Journal of Social Robotics* 13, 1 (Feb. 2021), 31–40. <https://doi.org/10.1007/s12369-020-00707-z>
- [14] Mark Coeckelbergh. 2022. *The political philosophy of AI: an introduction*. John Wiley & Sons.
- [15] European Commission. 2019. ETHICS GUIDELINES FOR TRUSTWORTHY AI.
- [16] Tara Patricia Cookson, Lorena Fuentes, Julia Margaret Zulver, and Austin Nelson. 2023. Fit for purpose? Assessing the accessibility, theory of action, and accountability of digital technology interventions for sexual and gender-based violence prevention and response. *Gender, Technology and Development* (2023), 1–23.
- [17] Kovila P.L. Coopamootoo and Magdalene Ng. 2023. "Un-Equal Online Safety?" A Gender Analysis of Security and Privacy Protection Advice and Behaviour Patterns. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 5611–5628. <https://www.usenix.org/conference/usenixsecurity23/presentation/coopamootoo>
- [18] Media Culture and Sport Committee. 2023. *Connected tech: smart or sinister?* Report. HC 2022-2023 157. <https://committees.parliament.uk/publications/41099/documents/200210/default/>
- [19] Kate Darling. 2012. Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects. <https://doi.org/10.2139/ssrn.2044797>
- [20] Maartje M. A. de Graaf, Somaya Ben Allouch, and Tineke Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior* 43 (Feb. 2015), 1–14. <https://doi.org/10.1016/j.chb.2014.10.030>
- [21] Anne P DePrince and Jennifer J Freyd. 2002. The intersection of gender and betrayal in trauma. In *Gender and PTSD*, R. Kimerling and P. Ouimette & J. Wolfe (Eds.). The Guilford Press, 98–113.
- [22] Carlos Bermejo Fernandez and Pan Hui. 2022. Life, the metaverse and everything: An overview of privacy, ethics, and governance in metaverse. In *2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 272–277.
- [23] Office for National Statistics (ONS). 2022. *Domestic abuse in England and Wales overview: November 2022*. Statistical Bulletin. <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2022>
- [24] Jodi Forlizzi. 2007. How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. ACM, 129–136.
- [25] Hideki Garcia Goo, Katie Winkle, Tom Williams, and Megan K Strait. 2023. Victims and Observers: How Gender, Victimization Experience, and Biases Shape Perceptions of Robot Abuse. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2439–2446.
- [26] Faith Gordon, Alyce McGovern, Chrissy Thompson, and Mark A Wood. 2022. Beyond cybercrime: New perspectives on crime, harm and digital technologies. , i–viii pages.
- [27] Susann Hartmann, Heike Wiesner, and Andreas Wiesner-Steiner. 2007. Robotics and gender: The use of robotics for the empowerment of girls in the classroom. In *Gender designs IT: Construction and deconstruction of information society technology*. Springer, 175–188.
- [28] Nicola Henry, Asher Flynn, and Anastasia Powell. 2020. Technology-facilitated domestic and sexual violence: A review. *Violence Against Women* 26, 15–16 (2020), 1828–1854.
- [29] Nicola Henry and Anastasia Powell. 2015. Embodied harms: Gender, shame, and technology-facilitated sexual violence. *Violence Against Women* 21, 6 (2015), 758–779.
- [30] Travis Holland. 2021. "My battery is low and it's getting dark": The opportunity Rover's collective persona. *Persona Studies* 7, 1 (2021), 90–100.
- [31] Meghan E Hollis-Peel, Danielle M Reynald, Maud Van Bavel, Henk Elffers, and Brandon C Welsh. 2011. Guardianship for crime prevention: A critical review of the literature. *Crime, law and social change* 56 (2011), 53–70.
- [32] Adam O Horvath and Lester Luborsky. 1993. The role of the therapeutic alliance in psychotherapy. *Journal of consulting and clinical psychology* 61, 4 (1993), 561.
- [33] Shogo Ikari, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2020. Multiple-Robot Mediated Discussion System to support group discussion. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 495–502.
- [34] Sooyeon Jeong, Laura Aymerich-Franch, Sharifa Alghowinem, Rosalind W Picard, Cynthia L Breazeal, and Hae Won Park. 2023. A robotic companion for psychological well-being: A long-term investigation of companionship and therapeutic alliance. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 485–494.
- [35] Sooyeon Jeong, Laura Aymerich-Franch, Sharifa Alghowinem, Rosalind W. Picard, Cynthia L. Breazeal, and Hae Won Park. 2023. A Robotic Companion for Psychological Well-being: A Long-term Investigation of Companionship and Therapeutic Alliance. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 485–494. <https://doi.org/10.1145/3568162.3578625>
- [36] Sooyeon Jeong, Laura Aymerich-Franch, Kika Arias, Sharifa Alghowinem, Agata Lapedriza, Rosalind Picard, Hae Won Park, and Cynthia Breazeal. 2023. Deploying a robotic positive psychology coach to improve college students' psychological well-being. *User Modeling and User-Adapted Interaction* 33, 2 (2023), 571–615.
- [37] Richard Johnson. 2002. Changing attitudes about domestic violence. *Law and Order* 50, 4 (2002), 60–65.
- [38] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. Association for Computing Machinery, New York, NY, USA, 229–236. <https://doi.org/10.1145/2696454.2696460>
- [39] Liz Kelly. 2013. *Surviving sexual violence*. John Wiley & Sons.
- [40] Cory D Kidd and Cynthia Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE, 3230–3235.
- [41] Abbas Z Kouzani. 2023. Technological innovations for tackling domestic violence. *IEEE Access* (2023).
- [42] Alyssa Kubota, Dagoberto Cruz-Sandoval, Soyon Kim, Elizabeth W. Twamley, and Laurel D. Riek. 2022. Cognitively Assistive Robots at Home: HRI Design Patterns for Translational Science. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 53–62.
- [43] Michael J Lambert and Dean E Barley. 2001. Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, research, practice, training* 38, 4 (2001), 357.
- [44] Michal Luria, Ophir Sheriff, Marian Boo, Jodi Forlizzi, and Amit Zoran. 2020. Destruction, catharsis, and emotional release in human-robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)* 9, 4 (2020), 1–19.
- [45] Francois Michaud, Patrick Boissy, Daniel Labonte, Helene Corriveau, Andrew Grant, Michel Lauria, Richard Cloutier, Marc-Andre Roux, Daniel Iannuzzi, and Marie-Pier Royer. 2007. Telepresence Robot for Home Care Assistance.. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*. California, USA, 50–55.
- [46] Lux Miranda, Ginevra Castellano, and Katie Winkle. 2023. Examining the State of Robot Identity. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 658–662.
- [47] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.

- [48] Natasha Mulvihill, Nadia Aghtaie, Andrea Matolcsi, and Marianne Hester. 2022. UK victim-survivor experiences of intimate partner spiritual abuse and religious coercive control and implications for practice. *Criminology & Criminal Justice* (2022), 17488958221112057.
- [49] BBC News. 2016. Dementia doll abuse carers at Rochdale nursing home detained. <https://www.bbc.com/news/uk-england-manchester-38050607>
- [50] Yohei Noguchi, Hiroko Kamide, and Fumihide Tanaka. 2020. Personality traits for a social mediator robot encouraging elderly self-disclosure on loss experiences. *ACM Transactions on Human-Robot Interaction (THRI)* 9, 3 (2020), 1–24.
- [51] Lesley Nuttall. [n.d.]. The Language We Use. <https://www.ibm.com/policy/wp-content/uploads/2020/06/Design-Principles-to-Combat-Domestic-Abuse.pdf>. Accessed: September 27, 2023.
- [52] Council of Europe. 2014. The Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence. <https://www.coe.int/en/web/istanbul-convention/key-facts>. Accessed: September 27, 2023.
- [53] World Health Organization et al. 2021. *Violence against women prevalence estimates, 2018: Global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women. Executive summary*. World Health Organization.
- [54] Caroline Criado Perez. 2022. 3. Computer says no – is AI making healthcare worse for women? <https://www.tortoisemedia.com/audio/visible-women-caroline-criado-perez-episode-3/>
- [55] Leonie Pihama, Ngaropi Cameron, and Rihi Te Nana. 2019. Historical trauma and whānau violence. (2019). <https://apo.org.au/node/263601>
- [56] New Zealand Police. 2023. *Annual Report 2021/22*. Report. <https://www.police.govt.nz/sites/default/files/publications/annual-report-2021-2022.pdf>
- [57] Anastasia Powell and Nicola Henry. 2017. *Sexual violence in a digital age*. Springer.
- [58] Helena Pozniak. 2023. Can tech help end sexual violence? <https://eandt.theiet.org/content/articles/2023/07/can-tech-help-end-sexual-violence/>
- [59] Linda Regan, Liz Kelly, Anne Morris, and Rebecca Dibb. 2007. If only we'd known': An exploratory study of seven intimate partner homicides in Englethorpe. *Child and Woman Abuse Studies Unit* (2007).
- [60] Kathleen Richardson and Charlotta Odland. 2023. *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*. Cham: Palgrave Macmillan.
- [61] Astrid M Rosenthal-von der Pütten, Nicole C Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C Eimler. 2013. An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics* 5 (2013), 17–34.
- [62] Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C. Eimler. 2013. An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5, 1 (Jan. 2013), 17–34. <https://doi.org/10.1007/s12369-012-0173-8>
- [63] Joe Saunders, Dag Sverre Syrdal, Kheng Lee Koay, Nathan Burke, and Kerstin Dautenhahn. 2015. "teach me—show me"—end-user personalization of a smart home and companion robot. *IEEE Transactions on Human-Machine Systems* 46, 1 (2015), 27–40.
- [64] Amanda Sharkey and Noel Sharkey. 2011. Children, the elderly, and interactive robots. *IEEE Robotics & Automation Magazine* 18, 1 (2011), 32–38.
- [65] Robert Sparrow. 2017. Robots, Rape, and Representation. *International Journal of Social Robotics* 9, 4 (Sept. 2017), 465–477. <https://doi.org/10.1007/s12369-017-0413-z>
- [66] Evan Stark. 2009. *Coercive control: The entrapment of women in personal life*. Oxford University Press.
- [67] Evan Stark. 2023. *The Coercive Control of Children*. Oxford University Press.
- [68] Francesca Stevens, Leonie M Tanczer, Frances Ridout, and Shane D Johnson. 2021. The Applicability of the UK Computer Misuse Act 1990 onto Cases of Technology-Facilitated Domestic Violence and Abuse. *London: UK Home Office/University College London. Publication forthcoming* (2021). https://www.ucl.ac.uk/computer-science/sites/computer_science/files/the_applicability_of_the_uk_computer_misuse_act_1990_onto_cases_of_technology_facilitated_domestic_violence_and_abuse.pdf
- [69] Litska Strikwerda. 2015. Present and future instances of virtual rape in light of three categories of legal philosophical theories on rape. *Philosophy & Technology* 28 (2015), 491–510.
- [70] Dina Utami, Timothy W Bickmore, and Louis J Kruger. 2017. A robotic couples counselor for promoting positive communication. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 248–255.
- [71] Stefan Vogler, Rachel Kappel, and Elizabeth Mumford. 2023. Experiences of technology-facilitated abuse among sexual and gender minorities. *Journal of interpersonal violence* (2023), 08862605231179724.
- [72] Mary Wakeham. 2021. *Animal abuse as a strategy of coercive control*. Ph.D. Dissertation. University of Bristol.
- [73] Sarah-Jane Walker, Marianne Hester, and Elizabeth McCarthy. 2023. The use of chemical control within coercive controlling intimate partner violence and abuse. *Violence Against Women* (2023), 10778012231197579.
- [74] David Wall. 2003. *Crime and the Internet*. Routledge.
- [75] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. *I'd blush if I could: closing gender divides in digital skills through education*. Technical Report. <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>
- [76] Brenda K Wiederhold. 2022. Sexual harassment in the Metaverse. *Cyberpsychology, Behavior, and Social Networking* 25, 8 (2022), 479–480.
- [77] Randi Williams, Safinah Ali, Nisha Devasia, Daniella DiPaola, Jenna Hong, Stephen P Kaputso, Brian Jordan, and Cynthia Breazeal. 2023. AI+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education* 33, 2 (2023), 325–383.
- [78] Tom Williams and Kerstin Sophie Haring. 2023. No Justice, No Robots: From the Dispositions of Policing to an Abolitionist Robotics. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 566–575.
- [79] Katie Winkle, P. Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2021. Assessing and Addressing Ethical Risk from Anthropomorphism and Deception in Socially Assistive Robots. In *2021 16th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.
- [80] Katie Winkle, Ryan Blake Jackson, Alexandra Bejarano, and Tom Williams. 2021. On the flexibility of robot social identity performance: benefits, ethical risks and open research questions for HRI. In *HRI Workshop on Robo-Identity*.
- [81] Katie Winkle, Ryan Blake Jackson, Gaspar Isaac Melsión, Dražen Bršćić, Iolanda Leite, and Tom Williams. 2022. Norm-Breaking Responses to Sexist Abuse: A Cross-Cultural Human Robot Interaction Study. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 120–129. <https://doi.org/10.1109/HRI53351.2022.9889389>
- [82] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 72–82. <https://doi.org/10.1145/3568162.3576973>
- [83] Katie Winkle, Gaspar Isaac Melsión, Donald McMillan, and Iolanda Leite. 2021. Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. ACM, New York, NY, USA.
- [84] Marty J Wolf, K Miller, and Frances S Grodzinsky. 2017. Why we should have seen that coming: comments on Microsoft's "taylor" experiment," and wider implications. *Acm Sigcas Computers and Society* 47, 3 (2017), 54–64.
- [85] Delanie Woodlock. 2017. The abuse of technology in domestic violence and stalking. *Violence Against Women* 23, 5 (2017), 584–602.
- [86] Delanie Woodlock, Mandy McKenzie, Deborah Western, and Bridget Harris. 2020. Technology as a weapon in domestic violence: Responding to digital coercive control. *Australian social work* 73, 3 (2020), 368–380.
- [87] Lillian Yang and Carman Neustaedter. 2020. An autobiographical design study of a long distance relationship: When telepresence robots meet smart home tools. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 129–140.
- [88] Elizabeth Yardley. 2021. Technology-facilitated domestic abuse in political economy: A new theoretical framework. *Violence Against Women* 27, 10 (2021), 1479–1498.
- [89] Garry Young. 2014. *Ethics in the virtual world: The morality and psychology of gaming*. Routledge.