Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

A principlist framework for cybersecurity ethics

Paul Formosa[a,*], *Michael Wilson*[b], *Deborah Richards*[c]

[a] *Department of Philosophy, Macquarie University, NSW 2109, Australia*
[b] *School of Law, Murdoch University, WA 6150, Australia*
[c] *Department of Computing, Macquarie University, NSW 2109, Australia*

ARTICLE INFO

Article history:
Received 29 October 2020
Revised 3 June 2021
Accepted 19 June 2021
Available online 25 June 2021

Keywords:
Cybersecurity ethics
AI ethics
Principlism
Privacy
Penetration testing
DDoS attacks
Ransomware

ABSTRACT

The ethical issues raised by cybersecurity practices and technologies are of critical importance. However, there is disagreement about what is the best ethical framework for understanding those issues. In this paper we seek to address this shortcoming through the introduction of a principlist ethical framework for cybersecurity that builds on existing work in adjacent fields of applied ethics, bioethics, and AI ethics. By redeploying the AI4People framework, we develop a domain-relevant specification of five ethical principles in cybersecurity: beneficence, non-maleficence, autonomy, justice, and explicability. We then illustrate the advantages of this principlist framework by examining the ethical issues raised by four common cybersecurity contexts: penetration testing, distributed denial of service attacks (DDoS), ransomware, and system administration. These case analyses demonstrate the utility of this principlist framework as a basis for understanding cybersecurity ethics and for cultivating the ethical expertise and ethical sensitivity of cybersecurity professionals and other stakeholders.

1. Introduction

The social and financial importance of cybersecurity is increasingly being recognised by governments. This includes, for the US alone, the roughly $100 billion annual cost of cyberattacks (Bouveret, 2018) and a corresponding cybersecurity market estimated to be worth $170 billion a year (Awojana and Chou, 2019). While there is much discussion around technical solutions to cybersecurity issues, there is far less focus on the ethical issues raised by cybersecurity. Cybersecurity is of critical ethical significance because cybersecurity technologies have an important impact on human well-being as they make possible many contemporary human organisations which rely on the accessibility and integrity of data and computer systems. Cybersecurity raises important ethical trade-offs and complex moral issues, such as whether to pay hackers to access data encrypted by ransomware or to intentionally deceive people through social engineering while undertaking penetration testing. However, when ethical issues in cybersecurity are explicitly discussed (e.g., Christen et al., 2020; Himma, 2007; Manjikian, 2018), there remains broad disagreement about the best conceptual framework for understanding these issues. To deal with this problem, we redeploy to a cybersecurity context a principlist framework, based upon literature in ethical AI (artificial intelligence) and bioethics (Floridi et al., 2018; Beauchamp and Childress, 2001), that focuses on five ethical principles: beneficence, non-maleficence, autonomy, justice, and explicability. These principles can conflict with one another and need to be balanced in a context-sensitive manner, which can result in a range of ethical trade-offs that we explore by examining the ethical issues raised by four common cybersecurity contexts: penetration (pen) testing, distributed denial of service attacks (DDoS), ransomware, and system ad-

* Corresponding author.
E-mail address: Paul.Formosa@mq.edu.au (P. Formosa).
https://doi.org/10.1016/j.cose.2021.102382

ministration. By focusing on these common cases, these analyses demonstrate the utility of this principlist framework as a basis for understanding cybersecurity ethics and for cultivating the ethical expertise of cybersecurity professionals and other stakeholders.

2. Approaches to cybersecurity ethics

There is an emerging literature on ethical issues related to cybersecurity or computer and information security (Nissenbaum, 2005). Cybersecurity is an academic discipline and profession organised around the pursuit of the security of data, networks, and computer systems (Manjikian, 2018; Brey, 2007). Various cybersecurity technologies, such as firewalls and encryption, are used to achieve this goal in the face of various threats, such as viruses or phishing attacks. Since most important "human institutions/practices … *rely upon* … the integrity, functionality, and reliability …. of data, systems, and networks" that cybersecurity technologies make possible, it follows that "ethical issues are at the core of cybersecurity practices" because these secure "the ability of human individuals and groups to live well" (Vallor, 2018, p. 4). While much of the academic literature specifically on the ethics of cybersecurity is fairly recent (Brey, 2007; Christen et al., 2020; Himma, 2007, 2008; Manjikian, 2018; Tavani, 2007; Timmers, 2019), this literature builds on earlier work on the "hacker ethic" (Himma, 2008, p. 205) and pioneering work in computer ethics more generally (e.g. Weizenbaum, 1972; for a brief history see Manjikian, 2018, pp. 16–20).

An important distinction is often made in this literature between the ethical issues related to state or national cybersecurity and those related to civil or commercial cybersecurity. The former grouping includes issues such as cyberwarfare and state-sponsored cyber-surveillance, which are typically discussed in terms of just war theory, state sovereignty, international relations, and national security (Meyer, 2020; Schlehahn, 2020). In contrast, the latter includes issues such as the hacking of commercial entities or end users, which are not typically discussed in such terms and which raise a different set of ethical issues (Nissenbaum, 2005; 2011). To focus our discussion, we limit ourselves to the ethical issues raised by computer or information security in the context of civil or commercial cybersecurity where issues such as just war theory, state sovereignty and national security are not central. We thus omit discussion of state cybersecurity cases (but for a treatment of such cases see: Efrony and Shany, 2018; Macnish, 2018; Manjikian, 2018; Meyer, 2020; Schlehahn, 2020; Stevens, 2020). Cybersecurity within a civil or commercial context encompasses not only threats to infrastructure that stores commercial or user data, but also an ancillary set of financial, psychological, and social harms associated with the everyday use of information and communications technologies. For example, decisions to regulate speech hosted on platforms (whether manually by a web administrator or as automated by a software engineer) involves applying normative standards of "fighting words" or "hate speech" that recognise threats to other users' psychological wellbeing (Goldenziel and Cheema, 2019; Klein, 2019). Such normative questions can arise during the everyday responsibilities of software engineers or system administrators who act as agents of users' cybersecurity.

Much of the discussion in the literature on cybersecurity ethics focuses on the conflict between privacy and security (Van de Poel, 2020). However, this focus is too limiting (Christen et al., 2020). First, because cybersecurity technologies are both a prerequisite for *ensuring* privacy (Zajko, 2018) and a means of *violating* privacy (Hildebrandt, 2013). Second, because (as we shall see below) there are many other relevant ethical considerations, which means that a focus on privacy alone is both too narrow and masks other important ethical issues. Privacy is not the *only* and not always the most *important* ethical concern in cybersecurity. Nonetheless, privacy remains of core importance to cybersecurity ethics as our below framework makes clear, although we argue that it needs to be anchored across a broader moral framework.

Two broad approaches to cybersecurity ethics have emerged. The first approach is to apply core underlying moral theories, such as utilitarianism, directly to cybersecurity issues. The second approach is to develop a cluster of mid-level ethical principles for cybersecurity contexts. In addition, both approaches make use of casuistry, which is a detailed case by case method of analysis (Kuczewski, 1998). These two approaches are common in other areas of applied ethics, such as bioethics or ethical AI (Beauchamp and DeGrazia, 2004).

In terms of the first approach, a common method is to directly apply the big three ethical theories of consequentialism, deontological (which usually means Kantian) ethics, and virtue ethics (e.g., Mouton et al., 2015). A prominent recent example of this approach in the cybersecurity ethics context is Manjikian (2018). However, there are several problems with this approach. First, by necessity, such an analysis tends to be overly simplistic. Each major ethical theory is complicated and there are competing versions and entrenched disagreements within each theory (Formosa, 2017). Applying these base moral theories to complicated real-world issues tends to skip over these details and ignores important disagreements in the literature that can lead to conflicting outcomes. Second, which of the big three ethical theories should we apply when they give conflicting results, given the persistent lack of agreement about which ethical theory (if any) is best? This is a problem because no theory has overwhelming normative authority (Beauchamp and DeGrazia, 2004). Third, it is far from straightforward how to get from abstract and general underlying moral theories to concrete cases in cybersecurity. For example, how do we get from the dignity of humanity to the ethics of hacking back against a foreign actor to prevent a DDoS attack which could impact innocent third parties? Fourth, this approach fails to clearly bring forth the ethical issues, values and principles that are most relevant in the specific domain in question. For example, even if we know that maximising utility, respecting humanity, or acting virtuously is most important, how do we get from that to the ethical minutiae of cybersecurity practices? While none of these well-known issues are fatal for this approach, they all remain significant problems to be overcome. The last two issues are particularly problematic in the context of developing a useful framework for cultivating the ethical sensitivity of cybersecurity professionals, since focusing on general ethical theories

fails to foreground the specific ethical issues at play in cybersecurity.

The second broad approach is to outline a series of mid-level and domain-specific principles. This approach is commonly known as "principlism" and it remains the "dominant approach in biomedical ethics today" (Shea, 2020b, p. 442), where the four basic principles of beneficence, non-maleficence, autonomy, and justice are widely used (Beauchamp and Childress, 2001). Rather than rely on any single general moral theory, these principles are affirmed from a range of different moral theories and common-sense moral intuitions (Shea, 2020b). These mid-level principles operate at a less general level than moral theories, while being explicitly connected to a particular normative domain such as bioethics. This approach raises its own set of problems, with the two most prominent being: 1) how to apply these mid-level principles to specific cases; and 2) how to deal with conflicts and tensions between these mid-level principles (Davis, 1995). The solution to the first problem is to provide "specification": the "principles must be specified to suit the needs of particular contexts" (Beauchamp and DeGrazia, 2004, p. 61). The solution to the second problem is to draw on casuistry and case analysis to demonstrate how the "balancing" of principles in concrete cases is achieved (Beauchamp and DeGrazia, 2004, p. 61). There is also debate about how many principles there should be and how those principles are justified (Davis, 1995; Shea, 2020b).

While both approaches have their strengths and weaknesses, we adopt a principlist approach here for two key reasons. First, because it is by far the most common approach to cybersecurity ethics, and it is also the most common approach in other areas of applied ethics, such as bioethics (Shea, 2020a) and AI ethics (Floridi and Cowls, 2019; Hagendorff, 2020). This allows us to build on a rich and widely appealing foundation. Second, because this approach is the most useful one for explicitly bringing forth both the relevant ethical principles in a particular domain (i.e., specification) and the ethical conflicts that exist through case analysis (i.e., balancing). This is particularly important when considering the use of an ethical framework in an education or training context. In such cases it is important to focus on the four components of ethical expertise (Rest et al., 1999). Specifically, these are focus (prioritising morality), sensitivity (recognising morality), judgement (deciding what morality requires), and action (doing what morality requires), as outlined in the "Morality Play" framework for developing ethical expertise (Staines et al., 2019).

Being aware of the different ethical principles at play in a specific domain is important for ethical *sensitivity* training as it helps in making the relevant ethical issues explicit so that they can be recognised in practice; making the ethical principles explicit is also important for training moral *focus* as it brings home the ethical importance of choices; focusing on balancing conflicts between principles can help us to generate concrete scenarios for training moral *judgement*; and showing how to resolve ethical conflicts in real-world cases can help to demonstrate moral *action*. This illustrates the potential utility of the framework as a useful basis for cultivating the ethical expertise of cybersecurity professionals. The necessity of incorporating "ethical reasoning development into engineering professional preparation" has been previously recognised, and this similarly applies to analogous technical skillsets such as cybersecurity (Hess et al., 2019, p. 83). This development requires the ability to reason with ethical principles and goes beyond knowing relevant codes of conduct (which we discuss below), since practitioners need to be able to deal with ethical "grey areas", conflicts, vagueness and incompleteness in ethical guidelines, and novel situations raised by new technologies (Hess et al., 2019, p. 83). However, we focus here primarily on *sensitivity* as we emphasise the importance of recognising ethical conflicts between principles, rather than arguing how to resolve those conflicts (i.e., judgement), since our goal here is to demonstrate the usefulness of a principlist framework for cultivating ethical sensitivity rather than resolve controversial substantive disagreements about specific cases.

3. Specifying a principlist framework for cybersecurity ethics

Most existing frameworks in cybersecurity ethics adopt a principlist approach. However, there are a large range of different sets of principles that might be developed and applied, and these sometimes overlap or conflict. Many of these frameworks are outlined in a recent edited volume that is an output from the CANVAS (Constructing an Alliance for Value-driven Cybersecurity) project (Christen et al., 2020). A summary of existing principlist frameworks in this area can be found within Table 1. While there is some overlap, the variety of principles demonstrates a lack of consensus about the best framework for understanding cybersecurity ethics. It is also apparent that many frameworks succumb to the problem of principle proliferation, whereby new principles are added in an effort to capture the diversity of moral concerns relevant to a particular domain.

While these frameworks are an excellent place to start, it remains unclear which specific principles should be used. This is an important problem that is not easily resolved. One solution is to attempt to integrate the various frameworks into a novel one, although this would not necessarily resolve disagreements about which principles ought to prevail in a consolidated list. A second solution is to bring this newer area of research into closer alignment with established areas of applied ethics, such as bioethics, by redeploying existing principles while remaining sensitive to domain-specific issues. An important example of this second solution is found in the nearby area of research on ethical AI. The AI4People framework (Floridi et al., 2018) for ethical AI is a useful model to draw on in this regard, as it builds on Beauchamp and Childress's four basic ethical principles (autonomy, nonmaleficence, beneficence, and justice). However, rather than map them onto a new set of principles, as Weber and Kleine (2020) attempt to do, they instead provide domain specifications of these same four basic principles in an AI context. They also add an extra fifth principle, *explicability*, which incorporates both intelligibility and accountability, that emerges organically as significant in the AI context. This additional fifth principle is also needed in the domain of cybersecurity ethics because the intelligibility of, and accountability for, cybersecurity policies, practices and technologies are also significant ethical concerns in this domain.

Table 1 – Summary of principlist frameworks for cybersecurity ethics.

Source	Ethical Principles
Van de Poel (2020)	1) security, 2) privacy, 3) fairness, and 4) accountability.
The Menlo Report (2012)	1) respect for persons, 2) beneficence, and 3) justice
Loi and Christen (2020)	1) privacy, 2) data protection, 3) non-discrimination, 4) due process and free speech, and 5) physical integrity
Weber and Kleine (2020)	1) efficiency and quality of service, 2) privacy of information and confidentiality of communication, 3) usability of services, and 4) safety.
Morgan and Gordijn (2020)	1) privacy, 2) protection of data, 3) trust, 4) control, 5) accountability, 6) confidentiality, 7) responsibility on business to use ethical codes of conduct, 8) data integrity, 9) consent, 10) transparency, 11) availability, 12) accountability, 13) autonomy, 14) ownership, and 15) usability.

We will adopt a similar solution here by developing, for the first time, a domain specification of the five ethical principles from the AI4People framework in a cybersecurity context. This allows the framework to reformulate widely accepted principles and to connect to a long tradition of applied ethics research, while still allowing for organic domain-specific modifications to emerge (such as in the treatment of privacy, discussed below). To show that an ethical framework developed for one technological context, that of AI, applies to a different context, that of cybersecurity, we need to demonstrate the usefulness and breadth of the framework applied to cybersecurity. To do that we first outline a streamlined principlist framework (specification) in this section, before illustrating the framework through case analysis of important cybersecurity issues (balancing) in the next section. The first key benefit of this framework over alternatives outlined above is that it better coheres with principlist approaches in related areas of applied ethics, thereby allowing the theory to tap into a rich theoretical vein of literature, achieve broad acceptability, and avoid ad-hocness. The second key benefit is the effectiveness of the framework (as shown in the next section) in identifying the full range of ethical issues in common cybersecurity contexts, which is important for cybersecurity ethics training, while avoiding problematic principle proliferation.

According to the framework (see Fig. 1), we can specify the five basic principles of cybersecurity ethics as follows:

- **Beneficence:** Cybersecurity technologies should be used to benefit humans, promote human well-being, and make our lives better overall.
  Non-maleficence: Cybersecurity technologies should not be used to intentionally harm humans or to make our lives worse overall.
  Autonomy: Cybersecurity technologies should be used in ways that respect human autonomy. Humans should be able to make informed decisions for themselves about how that technology is used in their lives.
- **Justice:** Cybersecurity technologies should be used to promote fairness, equality, and impartiality. It should not be used to unfairly discriminate, undermine solidarity, or prevent equal access.
  Explicability: Cybersecurity technologies should be used in ways that are intelligible, transparent, and comprehensible, and it should also be clear who is accountable and responsible for its use.

While in a principlist framework all principles stand on an equal footing, each principle can have a different weight in different contexts. For example, in some cases autonomy may be the most important principle and override concerns to benefit people, but in other cases the weighting might be the inverse with beneficence being the more important principle. Balancing principles requires sensitivity to the full range of ethical issues covered by the five principles and the good judgement needed to discern the relative weight of each principle and to resolve any ethical trade-offs in that specific context. skilful balancing also requires awareness that different principles may be more or less salient in different contexts.

The principles as outlined above remain at a high-level of abstraction thus far. The task of specification is to fill in the domain-relevant details. We start to do that here in Fig. 1 which outlines various cybersecurity relevant ethical issues that emerge for each principle. We further the job of specification in the next section where we show how these principles and underlying ethical concerns emerge in four common cybersecurity contexts. However, we remain throughout the paper at the level of principle specification. Future work could involve the development of detailed guidelines that follow from the principles we outline here, but such guidelines or codes are additional and complementary to, and do not remove the need for, the principled-informed ethical reasoning that we demonstrate here. But first we briefly outline what is covered by each principle, before explaining the special role that privacy, which appears in different forms under multiple principles, has in our specification of this framework.

Non-maleficence: Cybersecurity practices focus on the availability, integrity, and confidentiality of data and systems (Brey, 2007). When data or systems are unavailable (e.g. through a DDoS attack), have their integrity compromised (e.g. through a hacker modifying files), and where confidentiality is not maintained (e.g. when patients' digital records are obtained improperly), then harm follows. Preventing these harms falls under the principle of non-maleficence. There are various forms these harms could take, including: privacy violations (e.g. when data confidentiality is breached); financial harms (e.g. loss of earning because a website is inaccessible [Christen et al., 2017]); physical harms (e.g. where a cyber breach leads to physical harm, such as with Stuxnet [Hildebrandt, 2013]); psychological harms (e.g. harms to mental health and well-being that can result from data breaches [Molitorisz, 2020]); system (e.g. costly repairs to a system) and data harms (e.g. data recovery or restore costs); and repu-

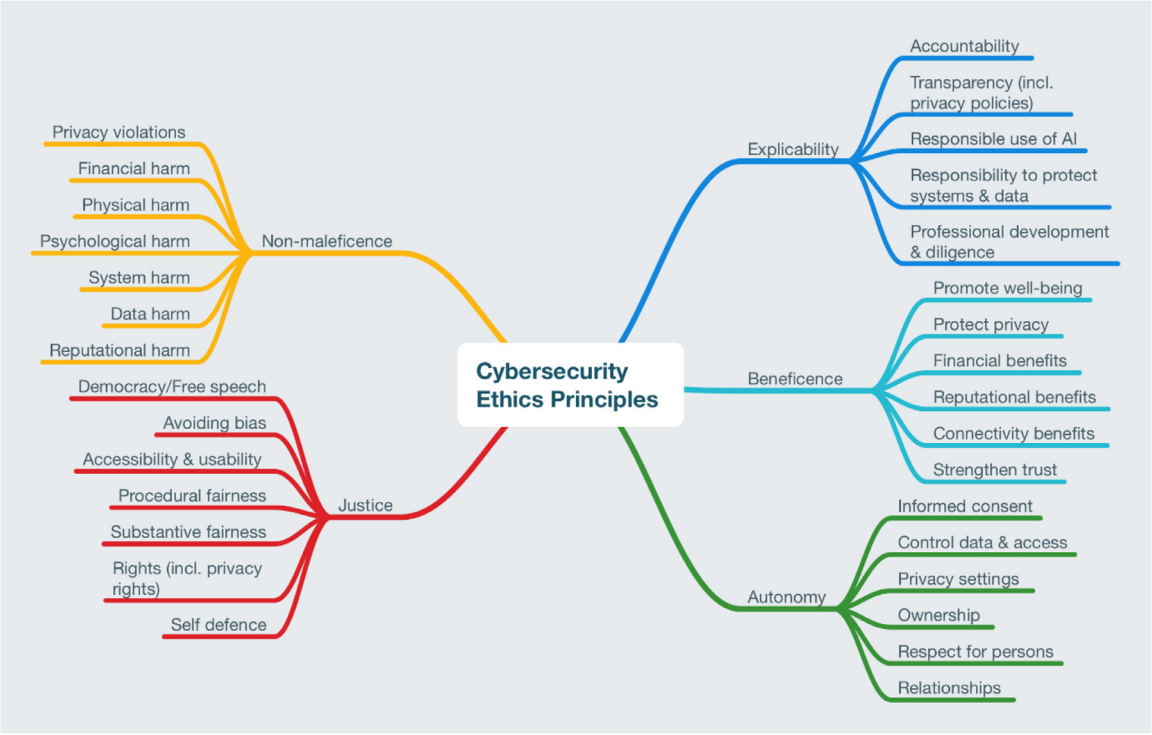

Fig. 1 – Five cybersecurity ethics principles.

tational harms (e.g. others won't trust your services if it is breached by hackers). The question of what *constitutes* harm, and how *severe* that harm is understood to be, is important for the application of the non-maleficence principle and its interaction with the other principles. We can see this, for example, in debates about what severity of "harm", such as a cybercrime, is necessary to justify another harm, such as privacy violations through covert surveillance (Simone, 2009; Harcourt, 1999).

Beneficence: Cybersecurity practices not only help us to avoid various harms, but they can also have many positive benefits and improve human well-being. Cybersecurity makes possible interactions, such as internet banking and e-commerce, that have enormous benefits (Manjikian, 2018). When we know that our systems and data are secured and accessible, we can interact with, store, and generate data with the confidence that it will be protected. There are many such benefits, including: promoting well-being (e.g. having personal data kept protected can be important for emotional health and well-being); protecting privacy (e.g. having one's privacy protected is important for self-development and negotiating relationships with others); financial benefits (e.g. good cybersecurity can have massive financial benefits [Awojana and Chou, 2019]); reputational benefits (e.g. improved reputation from having good cybersecurity could improve sales); connectivity (e.g. we can connect and share more openly with one another if the availability, integrity, and confidentiality of our data is ensured); and strengthen trust (e.g. we can develop trust in computer systems where good cybersecurity is in place). But as with non-maleficence, it matters what we understand as *counting* as benefits (e.g. do we focus only on financial benefits or also include harder to measure improvements in well-being?) and how we *quantify* different benefits (e.g. how do we weigh financial against well-being benefits?) when it comes to applying and balancing the beneficence principle.

Autonomy: Autonomy is about directing our own lives in accordance with our values (Formosa, 2013). This requires that we have control over who can access our data and systems. Consent is a key factor when it comes to autonomy since it is often through consent that we can rightfully obtain access to others and their data and systems (Molitorisz, 2020). Cybersecurity can both prevent unauthorised access to our data and facilitate access where consent is obtained. In terms of autonomy, cybersecurity can ensure: informed consent (e.g. cybersecurity can require us to get consent before accessing information and systems); control data and access (e.g. cybersecurity can give us control over our information and systems); privacy settings (e.g. users should have some control over their own privacy); ownership (e.g. to have a property right over our data requires that it be secured); respect for persons (e.g. treating people as ends in themselves means treating them as self-directing agents, which requires getting their consent when you wish to use them or their information [Formosa, 2017]); and relationships (e.g. relationships depend in part on being able to trust one another and share different types of information with each other which cybersecurity can help to make possible). This principle raises several important questions, such as whether we own all the data we generate, when a person's consent is needed to access or record their data, what counts as informed consent given the complexity of privacy settings (Nissenbaum, 2011), and when other considerations, such as preventing psychological harm to others (as part of

non-maleficence), can override the need to obtain an individual's consent.

Justice: Justice requires, amongst other things, ensuring fairness, accessibility and preventing bias (Floridi et al., 2018). In a cybersecurity context this includes: ensuring democracy and free speech (e.g. which requires accessibility of information and secure platforms for speech [Loi and Christen, 2020]); avoiding bias (e.g. ensuring cybersecurity technology, such as facial recognition in security cameras, is not biased against minorities [Hagendorff, 2020]); providing accessibility and usability (e.g. people need to be able to access and use information and systems, and some vulnerable groups may find it more difficult to navigate cybersecurity measures such as two factor authentication [Loi and Christen, 2020]); procedural fairness (e.g. following due process in dealing with a cybersecurity policy violation [Blanken-Webb et al., 2018]); substantive fairness (e.g. has a fair outcome been achieved?); protecting rights (e.g. are property, data and privacy rights being protected?); and allowing self-defence (e.g. to what extent can an organisation "hack back" against a DDoS attack? [Stevens, 2020; Tavani, 2007]). Justice also requires a focus on the distribution of harms and benefits and a consideration of their impacts on the least advantaged groups (Rawls, 1971) (e.g. does a choice of complicated login technologies prevent access to health records by elderly citizens who are most in need of these services?). Justice issues cover a broad range, and this can create internal tensions between different justice considerations. For example, a focus on accessibility of data and usability of systems for vulnerable users, such as by not requiring two factor authentication, can be in tension with ensuring the highest levels of cybersecurity to protect people's property rights in their data.

Explicability: Ethical cybersecurity systems and processes need to be explainable and transparent, and people and organisations need to be held accountable for their operation. This includes: accountability (e.g. who is responsible for a cybersecurity breach?); transparency (e.g. is it clear what cybersecurity policies and procedures are in place, including those around privacy?); the responsible use of AI in cybersecurity contexts (e.g. is the AI properly supervised and is it clear who is responsible for its operations? [Timmers, 2019]); and the responsibility of organisations and groups to protect systems and data (e.g. the responsibility to develop, maintain, and run good cybersecurity systems, policies, and practices). This last point emphasises the ethical importance of ongoing professional development and diligent work practices to ensure that the responsibilities of relevant computing professionals to implement, and keep updated, effective cybersecurity procedures and technologies is met. Explicability raises issues around what counts as best practice when it comes to cybersecurity, how to hold people and organisations accountable for failures of cybersecurity, and what levels of transparency are appropriate when it comes to cybersecurity operations.

Finally, given its importance in the cybersecurity ethics literature, we need to briefly justify the role of privacy in this framework, which is mentioned in Fig. 1 under each of the five principles rather than separated out as its own principle. Thomson (1975, p. 295) writes that "the most striking thing about the right to privacy is that nobody seems to have any very clear idea what it is." Privacy as a moral concept has been defined in a myriad of ways: a 'right to be let alone' derived from the principle of sovereign self-ownership and a right to exclude access to oneself (Warren and Brandeis, 1890, p. 205); a right to non-interference to prevent harms to oneself, such as the unauthorised access of private facts, publicising information in a false light, or appropriating one's identity (Prosser, 1960, pp. 390–401); an aspect of human dignity necessary to respect individuals as self-determining moral agents (Bloustein, 1964, p. 971); as a 'good' within the just society, necessary for establishing relationships characterised by "respect, love, friendship, and trust" (Fried, 1968, p. 475); and as a right to freedom from arbitrary surveillance as determined by a community of equals engaging in democratic deliberation (Newell, 2014, p. 521). As such, privacy is "a sweeping concept, encompassing (amongst other things) freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations" (Solove, 2008, p. 1). This multifaceted nature of privacy presents a challenge for articulating its role and function within a principlist framework.

Privacy could be incorporated within a principlist framework in several ways. For example, following Floridi et al.'s (2018, p. 697) AI4People framework, we could include privacy as a component of the principle of non-maleficence alone. While this approach clearly incorporates an understanding of privacy as a right to non-interference, it fails to accommodate the various other definitions of privacy within the philosophical literature noted above. To demonstrate this, Table 2 provides an outline of how the various definitions of privacy broadly relate to the other ethical principles within the AI4people framework.

It is important to note that the boundaries between these principles and the definitions of privacy are not absolute, and neither are the noted relationships one-to-one. For example, conceptualising privacy as a 'right to be let alone' as derived from the principle of sovereign self-ownership can also be linked with the principle of autonomy. The important point to note here is the problem of artificially narrowing the scope of 'privacy' to a single principle, such as non-maleficence.

An alternative approach to incorporating privacy is to include it as a separate ethical principle that attempts to incorporate the diversity of definitions noted above (e.g., Van de Poel, 2020; Loi and Christen, 2020; Morgan and Gordijn, 2020). While such an approach might more accurately reflect the multifaceted character of privacy, it perpetuates principle proliferation by ignoring the conceptual relationships between privacy and existing principles as summarised within Table 2. Consider, for example, the attempt to incorporate privacy as a unitary principle defined as 'freedom from unauthorised access to another individual's personal information'. But in articulating such a principle, we are still relying upon a more general principle of non-maleficence (i.e., where such unauthorised access is a type of harm that ought to be prevented). Similarly, such an approach complicates attempts to identify and define conflicts between 'privacy' and other ethical principles. For example, while there is an apparent conflict between 'privacy' and 'non-maleficence' where a system administrator is requested to provide another employee's personal information to assist with a criminal investigation,

Table 2 – Relationship between privacy and ethical principles.

Ethical Principle	Corresponding Definition of Privacy
Non-Maleficence	A right to non-interference to prevent harm (Prosser, 1960)
Justice	A 'right to be let alone' (Warren and Brandeis, 1890)
Explicability	A right to freedom from arbitrary surveillance (Newell, 2014)
Beneficence	A 'good' within the just society (Fried, 1968)
Autonomy	An aspect of human dignity (Bloustein, 1964)

this again relies upon a narrow conception of privacy as a right to non-interference. If we instead define privacy (with Newell, 2014) as 'freedom from *arbitrary* forms of surveillance', it may be reasonably claimed that 'privacy' has not been unduly violated here since the surveillance is not arbitrary. As such, a more accurate description of this conflict might be between 'autonomy' (respecting the privacy of persons by not accessing their personal information without consent), 'non-maleficence', and 'justice' (recognising that violating someone's privacy is a harm that may be necessary to prevent harm to others and achieve justice).

These various examples highlight the problem with conceptualising privacy as a single ethical concept (Solove, 2008, p. 9). Cognisant of these difficulties, rather than speak of 'privacy' as a unitary principle, we have instead subsumed it under the five more general ethical principles (see Fig. 1) so that we can more clearly identify relevant value conflicts. In doing so, we have organically modified previous principlist frameworks (where privacy has been either included under non-maleficence alone or separated out as a distinct principle) to better account for the multifaceted role of privacy within cybersecurity ethics.

4. Balancing ethical principles in cybersecurity

To continue the work of specifying and balancing the above five principles in a cybersecurity context, we shall engage in case analysis by exploring the following common cybersecurity scenarios: 1) penetration testing; 2) DDoS attacks; 3) ransomware; and 4) system administration. We picked these four cases as they represent scenarios that information and communications technology (ICT) professionals can regularly encounter. The Association for Computing Machinery (ACM) Code of Ethics (which we discuss further below) also provides several fictionalized scenarios "designed for educational purposes to illustrate applying the Code to complex situations" (ACM, 2018, p.13). Our inclusion of case studies in this article is driven by a similar goal. In presenting the case studies, we focus on demonstrating how the five principles outlined here can identify the full range of ethical issues that arise in common cybersecurity contexts by placing the relevant principle in brackets after identified ethical issues. Further, to ensure that the principlist framework is broadly applicable, we focus on the underlying ethical conflicts that exist regardless of jurisdictional or temporal differences in privacy law or computer crime statutes.

4.1. *Ethical issues in penetration testing*

The concept behind penetration (pen) testing, or "ethical hacking" (Martin, 2017), is that by using methods of bypassing security mechanisms that could be used by a nefarious actor, an organisation is able to identify and deal with vulnerabilities as part of cybersecurity risk mitigation. Pen testing can be undertaken internally within an organisation or externally by authorised cybersecurity firms (white hat), by unauthorised hackers seeking bug bounties and other rewards without intending to harm organisations (grey hat), and by hackers seeking to damage organisations and exploit vulnerabilities (black hat) (Manjikian, 2018). There can be clear benefits for customer security from exposing vulnerabilities that lead to fixes (beneficence), but if exposure of vulnerabilities occurs before a fix is available or if vulnerabilities are intentionally exploited then harm can result (non-maleficence). Pen testing can also violate natural property rights (justice), disrespect autonomy through the use of deception in social engineering (Hatfield, 2019), and lack transparency (explicability) depending on what agreements, if any, are in place beforehand. Organisations may also have responsibilities to undertake pen testing to ensure they have robust cybersecurity systems (explicability).

Two key ethical issues raised by pen testing are whether the pen tester has been authorised beforehand to undertake the cyberattack and how the hacker and relevant organisations deal with any vulnerabilities that are discovered. For example, Randal Schwartz, an Intel employee, was a system administrator who ran an unauthorised password crack which broke 48 of the 600 passwords he tested, including that of Intel's Vice President (Blanken-Webb et al., 2018). While there is little doubt Schwartz was acting in the interests of his organisation (see: Quarterman, 1995), the crack was reported by another Intel employee before Schwartz presented his findings to senior management. Consequently, Schwartz was accused of corporate espionage and the matter was referred to police for investigation. He was convicted in 1994 of three felonies broadly relating to the unauthorised access and modification of computer systems, although in 2007 his convictions were officially set aside (Leyden, 2007). Schwartz's case illustrates that pen testers risk violating an organisation's property and privacy rights (justice) and their autonomy if explicit authorisation is not obtained beforehand. Without transparency around his actions, Schwartz also risked violating the principle of explicability, even if his aims were to help his organisation (beneficence) without doing harm (non-maleficence). Bug bounty programs are another important case since they encourage grey hat hackers to undertake unauthorised pen

testing with the aim of discovering and reporting vulnerabilities (Manjikian, 2018). Such actions are ethically risky for grey hat hackers. For example, a 13-year-old Australian schoolboy who penetrated Apple's systems was motivated by a desire to impress the organisation to gain future employment with them. However, rather than land him a job, the schoolboy was charged with various computer hacking offences (Opie, 2019). This highlights the thin ethical line often faced by cybersecurity practitioners, since in other cases grey hat hackers have been financially rewarded rather than punished (Goodin, 2020). These examples show that, while beneficence and non-maleficence are important ethical goals for pen testing, autonomy, rights (or justice) and transparency (i.e., explicability) must also be respected.

These cases also raise the issue of how the "ethical hacker" should respond when vulnerabilities are detected. Martin (2017) contrasts an ethical "low road" of "immediate full [public] disclosure", which creates opportunities for black hat hackers to exploit exposed vulnerabilities, with an ethical "high road" of "responsible disclosure", which involves first disclosing vulnerabilities to impacted organisations privately and only publicly disclosing vulnerabilities after a fix or mitigation has been released. A complication occurs when an organisation fails to fix, or is unable to fix in a timely manner, a vulnerability after receiving notification from a pen tester. The pen tester then has the option of leaving the vulnerability publicly undisclosed, which leaves users unaware of a vulnerability that could be actively exploited, or disclosing publicly a vulnerability after a set period of time, which can lead to (or increase) its active exploitation in the absence of a fix. This case requires weighing up the benefits to users through disclosure of the vulnerabilities (beneficence), potential harm that both disclosure and nondisclosure may cause (non-maleficence), a requirement to be transparent (explicability), and the importance of meeting any contractual obligations that may be in place (justice and autonomy). The best ethical option will, as always, depend on the details of cases but if, for example, the benefits to users are very great (e.g. there are viable software alternatives) and the potential harms are very low (e.g. the chance of exploitation is not markedly increased by public disclosure), there are no prior contractual arrangements in place, and the process of detection was justifiable, then public disclosure may be appropriate, although this judgement won't apply to all cases.

There are several standard frameworks and methodologies for conducting penetration tests, including: the Open Source Security Testing Methodology Manual (OSSTMM), the Penetration Testing Execution Standard (PTES), the NIST Special Publication 800–115 (Scarfone et al., 2008), the Information System Security Assessment Framework, and the OWASP Testing Guide (for a comparison see Shanley and Johnstone, 2015). While standards aim to ensure that services and systems are safe and reliable, compliance is voluntary and the guidelines they provide are not associated with ethical principles (unlike ethical codes of conduct, which we explore below). This gap highlights that there is an important educative role for ethical frameworks such as the one presented here, as they can help to make the relevant ethical principles explicit and increase sensitivity amongst cybersecurity professionals to the range of ethical conflicts that can occur.

4.2. *Ethical issues in DDoS attacks*

A denial of service attack (DoS) is a cyberattack that attempts to deny access to a computer system or server (Mirkovic and Reiher, 2004). This attacks the availability of data or services, without undermining the confidentiality and integrity of data. This typically occurs by flooding a website with many more requests than it can handle (such as a HTTP request flood), making it difficult or impossible for legitimate users to access the site (Herrmann and Pridöhl, 2020). To create enough service requests to bring down a site, hackers often instigate distributed denial of service attacks (DDoS), which is a DoS attack that originates from multiple systems simultaneously and which usually involves using many hacked innocent third-party devices to send bogus requests to a server to undermine availability for legitimate users. This can involve using malware to infect other computers and devices to transform them into a botnet controlled by the attacker (Antonakakis et al., 2017). The use of botnets in a DDoS attack can, in comparison to a DoS attack, make it difficult to identify both the presence of an attack and the initiator of the attack, and include larger volumes of traffic and innocent third parties in any attack back scenarios.

There are two main types of responses to DDoS attacks (Dietzel et al., 2016; Himma, 2008; Martin, 2017): active responses (e.g. to attack back against the attacker) and passive responses (e.g. trying to block illegitimate traffic and increase bandwidth). There are ethical issues with both responses. The main difficulty with attempting to block the malicious traffic causing the denial of service is that the malicious traffic is often indistinguishable from legitimate traffic. One response to this is blackhole filtering which involves routing both legitimate and malicious traffic into a 'blackhole' where the request is dropped from the network (Dietzel et al., 2016). While this approach has benefits in keeping the site open for some users (beneficence), it comes at the cost of denying service to some legitimate traffic and thus harming innocent users (non-maleficence). This could be particularly significant if it involves access to important time-sensitive data, such as medical records. There are also justice concerns in the indiscriminate denial of service to some legitimate traffic. Further, the fact that the traffic has been routed to a 'blackhole' is not always made explicit to legitimate traffic and this can result in a lack of explanation as to the reason for the denial of service (explicability). Greater discrimination between legitimate and malicious traffic can help to offset some of these negative ethical consequences, as can the purchasing of more bandwidth, but more sophisticated DDoS attacks result in malicious traffic that is very difficult to detect and can overwhelm available bandwidth. Ethical solutions will seek to balance the need to minimise harms (e.g. by purchasing more bandwidth) with the requirement to be transparent (e.g. through announcements) and avoid bias and unfairness in blocking traffic (e.g. not blocking all traffic from, say, Africa).

Active responses to DDoS attacks involve "hacking back" (Himma, 2008). Himma (2008) differentiates between benign and aggressive responses. An aggressive response could involve attacking the attackers to try to prevent the denial of service. One version of this is to reroute the DoS attack pack-

ets back at the attackers to overwhelm their servers. An example of this is Conxion's response to a DoS attack by the Electrohippies (Himma, 2008). In contrast, an example of a benign response is to undertake a "traceback" to attempt to identify the perpetrators of the attack. Both responses raise ethical issues. An aggressive attack back will likely inflict harm on innocent third parties (non-maleficence). This could occur either through attacking innocent bots in a botnet or by negatively impacting innocent parties such as legitimate users who might be trying to access the sites being targeted by the attack back. A benign traceback, while it may not cause any significant harm, does involve unauthorised effects to victims' machines (autonomy), which may amount to an infringement of "the property rights of innocent person[s]" (justice) (Himma, 2008) and include privacy violations (non-maleficence). While active responses can cause unnecessary harm to others (non-maleficence), they also help to discourage future attacks and can help to end the current attack more quickly (beneficence). While tracebacks can help to facilitate justice by identifying the perpetrators, they can also undermine the property and privacy rights of, and fail to get consent from, impacted third parties (justice and autonomy). Further, insofar as such responses are often opaque and unexplained, they raise explicability concerns. Together these autonomy, justice and explicability concerns mean that the bar for ethically justifying attack backs on the grounds of avoiding harms or achieving benefits is not an easy one to meet.

Another issue is whether offensive DoS or DDoS attacks can be ethically justified. This occurs when the aim of the DoS attack is to achieve justice, benefit people, or frustrate a bad actor. This is also known as "hacktivism" (Efrony and Shany, 2018; Himma, 2008; Manjikian, 2018). One example is the 2006 case of German activists who carried out a DDoS attack against Lufthansa to "protest the fact that the airline was cooperating in the deportation of asylum seekers" (Manjikian, 2018). Another example is Operation Payback, which involved a DDoS attack on banks and payments sites, such as PayPal and Visa, that had withdrawn banking facilities from WikiLeaks (Mackey, 2010). Such cases raise several ethical issues. They are often intended to harm a powerful or disreputable group (non-maleficence) with the intent of helping another (often more vulnerable) group (beneficence). In the above German case, the attack was designed to harm Lufthansa and to help asylum seekers. But such attacks also harm innocent third parties (non-maleficence), such as customers wanting to purchase aeroplane tickets who are unable to do so since access to booking sites was denied by the DDoS attack. While the actions of hacktivists are typically non-violent, they do harm others (non-maleficence) and fail to get the consent of all impacted parties (autonomy), even if they aim to benefit others (beneficence). Some hacktivists also claim to be engaged in legitimate acts of civil disobedience aimed at changing unjust laws or policies (justice), although the legitimacy of this is strongly contested by others (Himma, 2008), since public explanation and the acceptance of legal responsibility are important components of civil disobedience (Rawls, 1971) and many hacktivists attempt to hide behind anonymity (Bodó, 2014). This suggests that ethical hacktivists should explain and accept legal responsibility for their actions (explicability) and seek to minimise harm to third parties.

A further issue is the decision of DDoS protection vendors, such as Cloudflare, to withdraw DDoS protection services to 8chan and related sites that host hate speech, incitements to violence, illegal content, or are connected to terrorist or racist attacks (Brodkin, 2020; for discussion of the Cloudflare and 8chan case see Taylor and Wong, 2019). Denying DDoS protection services to such sites opens those sites up to DDoS attacks by hacktivists, which risks harming some users of those sites (non-maleficence) and potentially restricting their users' free speech (justice), but can also help to prevent illegal activity, violence, terrorism, and hate speech (justice and non-maleficence). An organisation providing cybersecurity services also has a right (within bounds) to choose who they will provide protective services to (justice), and this reinforces the importance of both the provider and the recipient consenting to the provision of security services (autonomy). In this case, the autonomy of service providers to act on their values seems to override their obligation to protect others hosting questionable or illegal content, although there can also be legal obligations and rights at play in various jurisdictions, such as non-discrimination in service provision (justice), that may outweigh other ethical considerations.

4.3. *Ethical issues in ransomware attacks*

Ransomware attacks are becoming more common, with one study claiming that they had tripled between 2017 and 2018 (Morgan and Gordijn, 2020). WannaCry and Petya are two prominent recent ransomware attacks, with the former hitting the British NHS which resulted in cancelled medical appointments and diverted ambulances (Hern, 2017). Ransomware works by either encrypting data (cryptors) or blocking access to data (blockers) with the intention of extracting financial gains (Morgan and Gordijn, 2020). Typically, this outcome is achieved by offering to unencrypt or provide access to the data in exchange for the payment of a ransom. Although users sometimes gain access to their data after a ransom is paid, this is not always the case, meaning that the outcomes of paying a ransom are not clear (Herrmann and Pridöhl, 2020). This issue is complicated by the presence of cyberliability insurance cover, which creates a moral hazard by limiting the motives of organisations to prevent ransomware attacks as they will not have to bear the full costs of those breaches (Manjikian, 2018).

There are several responses to ransomware attacks that cybersecurity practitioners might pursue. First, try to isolate the damage by taking infected computers offline and then attempt to decrypt or gain access to the data. This has a low probability of success (Loi and Christen, 2020). Second, isolate the damage and then perform a full system and data recovery from unaffected backups. This assumes that up-to-date backups and the expertise to recover systems and data exists. There may also be significant system downtime while the recovery takes place, which can be costly (non-maleficence). Third, pay the ransom, or have one's insurer pay the ransom, and hope that the hacker provides access to the data after the payment is made. Further, some organisations have also chosen to attack the source of the ransomware to prevent pay-

ments. For example, the email box used by the authors of the Petya ransomware was deleted (Herrmann and Pridöhl, 2020), which meant users infected by the ransomware who wanted to pay for decryption keys (autonomy) could not contact the hackers. Another example is that of a government computer emergency response team (CERT) attacking ransomware by preventing access to payment servers so that victims cannot pay ransoms (Loi and Christen, 2020).

These various options each raise ethical issues. While blocking payment sites or services used by ransomware attackers might help to discourage future ransomware attacks (beneficence), it does so at the cost of harming those who need access to their encrypted data (non-maleficence) who no longer have the choice to pay hackers for that access (autonomy). This is particularly significant where the data is very important and where access is time sensitive, such as with medical records. Restoring data from backups can be time consuming and expensive, if it is even possible, and it can involve significant delays for access to systems and data which can in turn cause harms, such as missed medical appointments (non-maleficence). This can make restoring data more costly and more harmful than simply paying for decryption (Dudley, 2019). However, while paying for decryption might benefit users by giving them quick access to their data (beneficence), this can come at the significant cost of encouraging further ransomware attacks on others (non-maleficence). In terms of autonomy, attacking payment and email service providers used by ransomware attackers denies victims the choice of whether to pay a ransom to access their data. System operators may also have a legal obligation (justice) and a moral responsibility (explicability) to ensure they use best practices to protect and backup user data in their control (Fuster and Jasmontaite, 2020). This might involve putting in place protections to limit ransomware attacks, such as anti-phishing and social engineering training for staff, spam blockers for email systems, and data backup and recovery plans (Brewer, 2016). However, cyberliability insurance cover can complicate these matters, as it can make it cheaper to pay an insurance deductible in the event of a ransomware attack (beneficence) rather than pay to restore the system from backups or pay for better security to prevent the attack in the first place (justice) (Dudley, 2019). There can also be a lack of clear explanation regarding policies and practices that are in place to prevent and respond to ransomware attacks, as well as failures to hold to account those responsible for poor cybersecurity practices or a lack of professional development and diligence (explicability). Paying for insurance does not alleviate the ethical obligation to prevent ransomware attacks through investing in good security measures and implementing backup and recovery plans (explicability), and the choice of whether to pay a ransom must consider not only individual benefits but also the harms imposed on others through increasing the attractiveness of ransomware (non-maleficence).

4.4. *Ethical issues in cybersecurity system administration*

The system administrator role is important for ensuring the security of an organisation's computer systems. System administrators are typically responsible for giving users access to the internet and organisational IT resources in an equitable manner (justice), managing file servers (beneficence) and organisational firewalls (non-maleficence), monitoring internet connections and local area networks (LAN) for threats, and ensuring the latest security protocols and software are in place (non-maleficence and explicability). Decision making will often involve choosing settings and defaults (e.g. on servers and firewalls) that will have consequences on utilisation of ICT resources (beneficence) and deciding who has what level of access to ICT resources (autonomy) to minimise risk (non-maleficence). These decisions may restrict an individual's access to resources (justice) and their ICT choices (autonomy), and therefore these decisions should be transparent (explicability) without making the organisation vulnerable to cyberattacks (non-maleficence). Surveillance by system administrators of the ICT behaviour of users for the benefit (beneficence) and protection (non-maleficence) of the organisation poses an important privacy issue for individuals through the monitoring of their ICT usage (non-maleficence and autonomy).

System administrators face many dilemmas where the five ethical principles compete with one another. One collection of dilemmas involves decisions about how much agency end users should be given with regards to security and system updates and settings. For example, a system administrator might decide to use ethical worms to ensure that devices have up-to-date protection (non-maleficence), yet this conflicts with seeking the consent of device owners (autonomy) and respecting their ownership rights (justice) (Aycock and Maurushat, 2008). Automating updates raises questions about a fair distribution of the costs and benefits of ICTs (justice). Decisions to automate updates can disproportionately disrupt device usability amongst end users with disabilities if program or system interfaces are impacted (Vaniea and Rashidi, 2016; Gor and Aspinall, 2015). Automating security updates also impedes the visibility of such measures to end users (explicability), thereby depriving them of potential opportunities to learn basic cybersecurity skills (Wash et al., 2014). This is important because of the existence of a "digital divide" between social groups, which leads to different levels of exposure to vulnerabilities according to the underlying distribution of technical expertise (Dodel and Mesch, 2018; Albrechtsen and Hovden, 2009). Yet the problem cannot be resolved simply through the complete automation of security updates as human factors are an unavoidable part of cybersecurity. For example, as long as some updates (e.g. system critical updates) require human input to preserve the utility of devices, ensuring that end users are aware of the purpose, functions, and scope of security settings remains desirable (Vaniea and Rashidi, 2016) and helps to avoid security breaches (non-maleficence). Similarly, human factors are an unavoidable aspect of user authentication, including via password management, resistance to social engineering and phishing attacks, and anti-bot measures such as CAPTCHA tests (Hoonakker et al., 2009; von Ahn et al., 2003). Thus, system administrators must balance ethical concerns around the avoidance of harms through automating updates, with the limitations this places on users' autonomy, the complex justice considerations it raises in terms of usability and digital divides, and explicability issues about cybersecurity awareness through transparency.

System administrators must also consider questions about how the oversight of computer systems can influence interac-

tions between end users. The provision of cybersecurity, like other forms of security, can require regulation of users' behaviour to prevent harms, such as hate speech or reputational harms (Klein, 2019). At the most basic level, decisions must be made about what is acceptable user-generated content because the design and governance of online platforms structure how users interact with one another (for recent discussions of platform governance, see: Balkin, 2018; Roberts, 2018). Platform governance, and the ethical issues it raises, are thus an inevitable feature of providing platforms to users. For example, a user's right to free expression (autonomy and justice) often conflicts with, and may be limited by, other users' rights to not be harmed by hate speech (non-maleficence) (Balica, 2017; Banks, 2010). In turn, the degree to which restrictions on speech will be recognised as legitimate will also depend on the characteristics of the platform, with private platforms likely to attract stricter controls than public-facing websites, forums, or social media networks (Alonso, 2017; Mangan, 2018). By extension, when end users publish information on public-facing platforms using company property or in the course of their employment, decisions about acceptable speech will also turn on the importance of preventing reputational harms (non-maleficence) (Mangan, 2018). System administrators may also be charged with surveillance functions, documenting instances of inappropriate data access and monitoring employee performance (Lugaresi, 2010). System administrators must therefore balance competing interests in not harming end users through violating their privacy (non-maleficence), enhancing the accountability of users for their behaviour (explicability), and preventing social or financial harms to an organisation (non-maleficence). This balancing is further complicated by a risk that speech regulation may have a perverse consequence of silencing meaningful speech (autonomy) that benefits an organisation or community (beneficence and justice) (i.e. the "chilling effects" of digital surveillance as observed by Penney (2016)).

Finally, system administrators are often responsible for establishing and implementing an organisation's ICT policies and procedures, including end user codes of conduct (Wilk, 2016) and privacy policies, as part of achieving cybersecurity aims of secure data and system access. As such, system administrators are faced with decisions about the substantive contents of ICT policies, the suitable scope of consultation during policymaking processes (justice), and how to ensure compliance (explicability). For example, to ensure fairness and counteract bias, diverse representation across the organisation should be involved in the creation and review of such policies (justice), especially regarding ethically sensitive policies around privacy. Yet the exact weight that should be ascribed to different views remains contestable. The elicited preferences of management and end users may conflict with expert advice about preventing harm to an organisation (non-maleficence) or ensuring fairness in the enforcement of codes of conduct (*procedural* justice) (e.g. Shires, 2018; Cowley and Greitzer, 2015). Similarly, if too much weight is ascribed to the decisions of automated cybersecurity systems without adequate transparency (explicability), this may reduce perceptions of procedural legitimacy amongst end users (Danaher, 2016). System administrators thus need to ensure ICT policies are fair and arrived at through a just process (justice), are fit for purpose in preventing harm (non-maleficence) and benefiting users (beneficence), allow room for individual choice where appropriate (autonomy), and are transparent and justifiable (explicability).

5. Implications and limitations

The previous section has demonstrated that in common cybersecurity contexts there exists conflicts *between* and *within* different ethical principles (i.e. inter-principle and intra-principle conflicts). The presence of such principled ethical conflicts within the domain of civil and commercial cybersecurity practices highlights the importance of cultivating the ethical sensitivity of those who work with ICTs. Indeed, our consideration of comparatively mundane case studies, as opposed to matters of state cybersecurity, demonstrates how ethical decision-making is an unavoidable aspect of the everyday practices of cybersecurity and ICT professionals. Recognising the unavoidably normative character of such decisions is also important for illustrating the dangers of moral disengagement amongst those trained and employed in science and technology, where there is an observed tendency to adopt purely technocratic modes of decision-making (Cech, 2014; Grosz et al., 2019). However, as our analysis shows, there are no purely technocratic answers to many cybersecurity problems and ignoring ethical dilemmas does not make them disappear. For example, everyday decisions by system administrators to engage in pen testing or the use of ethical worms to minimise harm can undermine user autonomy, while simply displacing responsibility for cybersecurity onto users directly risks exacerbating problems of justice. The cybersecurity domain is thus fraught with ethical conflicts and trade-offs, problematising attempts to technocratically outsource decision-making to algorithms. Rather than attempt to resolve such conflicts here, which requires good judgement and depends on the specificity of cases, we have instead demonstrated how the five principles in our framework can expose the full range of these often-neglected ethical conflicts to sensitise practitioners to their presence.

Clearly, there is a need for ethical guidance in this area. Various IT professional societies, such as the ACM and the Institute of Electrical and Electronics Engineers (IEEE), provide a Code of Ethics and Professional Conduct for their members. We now demonstrate how our five ethical principles can be mapped on to these codes, which shows both how our framework can provide a principlist *underpinning* for such codes and how these codes can provide *complementary* details on top of our principles. For example, the IEEE (2020) Code of Ethics has a strong emphasis on professional behaviour and includes 10 principles. These include a focus on: non-maleficence (e.g. principle 1 on holding "paramount the safety, health, and welfare of the public…, [and] protect[ing] the privacy of others"); autonomy and justice (e.g. principle 7 on treating "all persons fairly and with respect" and not engaging in unjust "discrimination"); and explicability in the form of the responsibility to outline conflicts of interest (principle 3), engage in professional development and diligence (e.g. principle 5 on seeking and offering "honest criticism of technical work"), and support adherence to the code (principle 10). Overall, the focus

of this code is more on avoiding harms rather than on doing good (beneficence). Covering all computing professionals, the ACM (2018) Code of Ethics includes seven general ethical principles (numbered 1.1 to 1.7), nine professional responsibilities (numbered 2.1 to 2.9) that largely concern competent conduct of duties (2.1 - 2.6), seven professional leadership principles (numbered 3.1–3.7) and two principles for compliance with the code. These (ACM, 2018) map onto our principles as follows: non-maleficence (1.2 "Avoid harm"; 1.6 "Respect privacy"; 2.5 risk analysis; 2.9 "robustly and useably secure" systems; 3.6 & 3.7 "Use care" in changing and integrating systems), beneficence (1.1 "Contribute to society and to human well-being", 3.1 "Ensure public good", 3.2 "Enhance quality of working life"), justice (1.4 "Be fair" and do not "discriminate"; 1.7 "honour confidentiality"; 2.3 "Respect existing rules"), autonomy (1.5 "Respect the work" of others; 2.8 Authorised/essential access; 3.5 "Create opportunities"), and explicability (1.3 "Be honest and trustworthy" and be accountable and transparent; 2.7 "Foster public awareness"; 3.2 "Articulate social responsibilities"). However, it should be noted that certain clauses could map onto more than one principle, such as 2.9 "robustly and useably secure" systems which we have placed under non-maleficence given its primary focus on robust security to avoid harm but which could also belong under justice given its focus on usability as well.

More specific Codes of Ethics for cybersecurity professionals also exist. For example, the Information Systems Security Association (ISSA) outlines a Code of Ethics with six principles. These principles (ISSA, 2007) cover: justice (e.g. acting in "accordance with all applicable laws" and maintaining "appropriate confidentiality"), non-maleficence (e.g. not intentionally injuring the reputations of "colleagues, clients, or employers"), and explicability (e.g. avoiding conflicts of interest and discharging "professional responsibilities with diligence and honesty"). However, the code does not explicitly consider respecting (autonomy) and benefiting others (beneficence), and while the code requires acting in accordance with "the highest ethical principles" (ISSA, 2007), it does not provide guidance as to how conflicts between different ethical principles are to be balanced and how the code is to be applied in practice.

While such codes have various uses and provide important details (Shanley and Johnstone, 2015), the evidence of the effectiveness of ethical codes at improving moral behaviour is mixed, with some studies showing exposure to codes of conduct can reduce unethical decisions while other studies show no significant effect (McNamara et al., 2018, p. 730). In any case, such codes do not relieve cybersecurity professionals of the need to make informed ethical judgments of their own, or provide guidance on dealing with ethical "grey" areas or conflicts within the code ((Hess, 2019). To engage in independent ethical reasoning, cybersecurity professionals need to be aware of the ethical principles, such as those outlined here, that *underlie* more detailed ethical guidelines and codes of conduct and be able to make their own ethical judgments based on awareness of the relevant ethical principles in common scenarios (as outlined in Section 4). For example, through being sensitised to the underlying ethical principles an individual can evaluate the entire ACM Code of Conduct (including professional responsibilities and leadership) according to these principles, as was done above, and use similar reasoning to identify what ethical issues are raised when faced with a novel dilemma in practice. These principles can thus help to sensitise ICT professionals to the range of underlying ethical principles implicit in such codes.

The unavoidably ethical character of cybersecurity decision making highlights the importance of developing a normative framework that is suitable for the domain and has been developed specifically to help ensure that cybersecurity professionals become "aware that there is a moral problem when it exists" (Rest et al., 1999, p. 101). Indeed, in contrast to the high levels of abstraction required for the direct application of consequentialist, deontological, or virtue-orientated theories, principlist frameworks provide a more suitable foundation for the moral education of cybersecurity professionals accustomed to structured frameworks of problem-solving (Beever and Brightman, 2016). Such a structured approach to cybersecurity ethics allows for the systematic detection and naming of ethical conflicts, without impeding subsequent flexibility in forming context-sensitive ethical judgments. The framework thus provides a domain-orientated language for encouraging moral deliberation within cybersecurity training and educational contexts and highlights how the five ethical principles interact with one another in real-world contexts. These skills might be cultivated through cybersecurity ethics training programs embedded within organisation-based training or tertiary education curricula (Wilk, 2016). The use of serious games for ethical training is another promising avenue for cybersecurity ethics training (Hendrix et al., 2016; Staines et al., 2019; Richards et al., 2020).

There are five important limitations of our paper. First, we intentionally excluded the consideration of cases of international state cyberwarfare and state cybersurveillance. Indeed, there are unique ethical issues associated with how malicious state actors complicate the balance between physical and cyber security within a state and the privacy of citizens (Manjikian, 2018; Nissenbaum, 2005). However, these ethical issues are beyond the scope of most cybersecurity professionals working in the private sector, and therefore they were not considered here. Further research could extend our approach to include such issues. Second, since we have primarily focused here on the value of a principlist framework for cultivating ethical *sensitivity* within a cybersecurity context, future research is necessary to demonstrate the utility of the framework for also cultivating reasonable ethical *judgement* amongst cybersecurity professionals. Such research might examine how the framework can be applied in cybersecurity education and training contexts to assist professionals in resolving controversial cases by balancing competing principles. Third, the adoption of a principlist framework structured around domain-specific case studies, to the exclusion of general moral frameworks such as consequentialism, deontological ethics, and virtue ethics, should be recognised as (in part) a practical trade-off for pedagogical purposes (Beever and Brightman, 2016; Bulger, 2007). By their nature, the principles chosen for the framework (as with all principlist frameworks) are derived from common-sense intuitions that are more comprehensively elucidated by those more general moral theories (Beauchamp and Childress, 2001, p. 389). Future work could explore the derivation of our principles from those general theories. Fourth, even mid-level principles re-

tain a certain degree of abstraction. To address this, future work could involve the development of detailed guidelines that follow from the principles outlined here, although such guidelines do not remove the need for ethical sensitivity and principled ethical reasoning. Fifth, given the mixed evidence about the effectiveness of codes of conduct at improving ethical behaviour (McNamara et al., 2018, p. 730), the effectiveness of our principles for helping ICT professionals to recognise ethical issues and conflicts in cybersecurity contexts needs empirical verification.

6. Conclusion

While the financial importance of cybersecurity is becoming increasingly recognised, the important ethical issues that cybersecurity raises are less well understood. In this paper we have sought to address this shortcoming through the introduction of a principlist ethical framework for cybersecurity that builds on existing work in adjacent fields of applied ethics. The present framework involves the first domain-relevant specification of the five ethical principles of beneficence, non-maleficence, autonomy, justice, and explicability in a cybersecurity context. This principlist framework allows us to identify a range of inter-principle and intra-principle ethical conflicts in cybersecurity, while both avoiding principle proliferation and effectively integrating with principlist approaches widely used in related areas of applied ethics. We illustrated these ethical trade-offs through exploring four common cybersecurity scenarios: penetration testing, DDoS attacks, ransomware, and system administration. These examples help to map out the variety of ethical trade-offs that cybersecurity professionals can face in their work and demonstrates the usefulness of the framework as a basis for training aimed at improving the ethical sensitivity of cybersecurity professionals and other stakeholders.

Author statements

All authors contributed to the conceptualization, writing, and reviewing of this article. The authors are listed in order of the degree of contribution.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Paul Formosa: Conceptualization, Writing - original draft, Writing - review & editing. **Michael Wilson:** Conceptualization, Writing - original draft, Writing - review & editing. **Deborah Richards:** Conceptualization, Writing - original draft, Writing - review & editing.

REFERENCES

Association for Computing Machinery (ACM). ACM Code of Ethics and Professional Conduct; 2018 http://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-booklet.pdf.

Albrechtsen E, Hovden J. The information security digital divide between information security managers and users. Comput. Secur. 2009;28(6):476–90.

Alonso DA. Social media in the employment relationship Context. Comp. Labor Law Policy J. 2017;39:287.

Antonakakis M, et al. Understanding the Mirai botnet. Proceedings of the 26th USENIX Security Symposium, 2017.

Awojana T, Chou T-S. Overview of learning cybersecurity through game based systems. Proceedings of the 2019 Conference for Industry and Education Collaboration, 2019.

Aycock J, Maurushat A. Good" worms and human rights. ACM SIGCAS Comput. Soc. 2008;38(1):28–39.

Balica R. The criminalization of online hate speech. Contemp. Read. Law Soc. Justice 2017;9(2):184–90.

Banks J. Regulating hate speech online. Int. Rev. Law Comput. Technol. 2010;24(3):233–9.

Beauchamp TL, Childress JF. Principles of Biomedical Ethics. Oxford University Press; 2001.

Beauchamp TL, DeGrazia D. Principles and principlism. In: Khushf G, editor. In: Handbook of Bioethics. Springer; 2004. p. 55–74.

Beever J, Brightman AO. Reflexive principlism as an effective approach for developing ethical reasoning in engineering. Sci. Eng. Ethics 2016;22(1):275–91.

Blanken-Webb J, et al. In: 2018 USENIX Workshop on Advances in Security Education. A case study-based cybersecurity ethics curriculum. Baltimore; 2018.

Bloustein EJ. In: Privacy As an Aspect of Human Dignity, 39. New York University Law Review; 1964. p. 962.

Bodó B. Hacktivism 1-2-3. Internet Policy Rev. 2014;3(4):1–12.

Bouveret A. *Cyber Risk For the Financial Sector* (Working Paper WP/18/143). International Monetary Fund; 2018.

Brewer R. Ransomware attacks. *Netw. Sec., 2016* 2016(9):5–9.

Brey P. Ethical aspects of information security and privacy. In: Petković M, Jonker W, editors. In: Security, privacy, and Trust in Modern Data Management. Springer; 2007. p. 21–36.

Brodkin J. *QAnon/8chan Sites Back Online After Being Ousted By DDoS-protection vendor*. Ars Technica; 2020 https://arstechnica.com/tech-policy/2020/10/qanon-8chan-sites-back-online-after-being-ousted-by-ddos-protection-vendor/.

Bulger JW. Principlism. Teach. Ethics 2007;8(1):81–100.

Cech EA. Culture of disengagement in engineering education? *Science*. Technol. Hum. Values 2014;39(1):42–72.

Christen, M., Gordijn, B., & Loi, M. (Eds.). (2020). *The Ethics of Cybersecurity*. Springer.

Christen M, Gordijn B, Weber K, Van de Poel I, Yaghmaei E. A review of value-conflicts in cybersecurity. ORBIT 2017(1):1.

Cowley JA, Greitzer FL. Organizational impacts to cybersecurity expertise development and maintenance. Proc. Hum. Factors Ergonom. Soc. Annu. Meet. 2015;59(1):1187–91.

Danaher J. The threat of algocracy. Philos. Technol. 2016;29(3):245–68.

Davis RB. The principlism debate. J. Med. Philos. 1995;20(1):85–105.

Dietzel C, Feldmann A, King T. In: International Conference on Passive and Active Network Measurement. Blackholing at IXPs Heraklion; 2016.

Dodel M, Mesch G. Inequality in digital skills and the adoption of online safety behaviors. In: Inf., Commun. Soc., 21; 2018. p. 712–28.

Dudley R. The extortion economy. ProPublica 2019 https://www.propublica.org/article/

the-extortion-economy-how-insurance-companies-are-fueling-a-rise-in-ransomware-attacks.
Efrony D, Shany Y. A rule book on the shelf? Tallinn manual 2.0 on cyberoperations and subsequent state practice. Am. J. Int. Law 2018;112(4):583–657.
Floridi L, Cowls J. A unified framework of five principles for AI in society. Harv. Data Sci. Rev. 2019;1(1).
Floridi L, et al. AI4People—An ethical framework for a good AI society. Minds Mach. 2018;28(4):689–707.
Formosa P. Kant's conception of personal autonomy. J. Soc. Philos. 2013;44(3):193–212.
Formosa P. Kantian ethics, Dignity and Perfection. Cambridge University Press; 2017.
Fuster G, Jasmontaite L. Cybersecurity regulation in the European Union. In: The Ethics of Cybersecurity. Springer; 2020. p. 97–118.
Fried C. Privacy. Yale Law J. 1968;77(3):475–93.
Goldenziel JI, Cheema M. The new fighting words? Univ. Pa. J. Const. Law 2019;22(1):81–170.
Goodin D. Apple pays $288,000 to white-hat hackers who had run of company's network. Ars Tech. 2020 https://arstechnica.com/information-technology/2020/10/white-hat-hackers-who-had-control-of-internal-apple-network-get-288000-reward/.
Gor B, Aspinall D. In: *Symposium On Usable Privacy and Security* (SOUPS). Accessible banking. Ottawa; 2015.
Grosz BJ, et al. Embedded ethics. Commun. ACM 2019;62(8):54–61.
Hagendorff T. The ethics of AI Ethics. Minds Mach. 2020;30:99–120.
Hatfield J. Virtuous human hacking. Comput. Secur. 2019;83:354–66.
Harcourt BE. The collapse of the harm principle. J. Crim. Law Criminol. 1999;90(1):109–94.
Hendrix M, Al-Sherbaz A, Bloom V. Game based cyber security training. Int. J. Serious Games 2016;3(1):53–61.
Hern A. WannaCry, Petya, NotPetya. The Guardian; 2017 https://www.theguardian.com/technology/2017/dec/30/wannacry-petya-notpetya-ransomware.
Herrmann D, Pridöhl H. Basic concepts and models of cybersecurity. In: The Ethics of Cybersecurity. Springer; 2020. p. 11–44.
Hess JL, et al. Enhancing engineering students ethical reasoning. J. Eng. Educ. 2019;108(1):82–102.
Hildebrandt M. Balance or trade-off? Philos. Technol. 2013;26(4):357–79.
Himma KE. Internet Security. Jones & Bartlett Learning; 2007.
Himma KE. Ethical issues involving computer security. In: Himma KE, Tavani HT, editors. In: The Handbook of Information and Computer Ethics. Wiley; 2008. p. 191–218.
Hoonakker P, Bornoe N, Carayon P. Password authentication from a human factors perspective. Proc. Hum. Factors Ergon. Soc. Ann. Meet. 2009;53(6):459–63.
Institute of Electrical and Electronics Engineers (IEEE). IEEE Code of Ethics 2020 https://www.ieee.org/about/corporate/governance/p7-8.html.
Information Systems Security Association (ISSA). ISSA Code of Ethics 2007 https://www.issa.org/issa-code-of-ethics/.
Klein A. From Twitter to Charlottesville. Int. J. Commun. 2019;13:297–318.
Kuczewski M. Casuistry and principlism. Theor. Med. Bioeth. 1998;19:509–24.
Leyden J. Intel "hacker" Clears his Name. The Register; 2007 https://www.theregister.com/2007/03/05/intel_hacker_charges_quashed/.
Loi M, Christen M. Ethical frameworks for cybersecurity. In: The Ethics of Cybersecurity. Springer; 2020. p. 73–96.
Lugaresi N. Electronic privacy in the workplace. Int. Rev. Law, Comput. Technol. 2010;24(2):163–73.
Mackey R. Operation Payback. The Lede; 2010 https://thelede.blogs.nytimes.com/2010/12/08/operation-payback-targets-mastercard-and-paypal-sites-to-avenge-wikileaks/.
Macnish K. Government surveillance and why defining privacy matters in a post-Snowden world. J. Appl. Philos. 2018;35(2):417–32.
Mangan D. Online speech and the workplace. Compar. Labor Law Policy J. 2018;39(2):357–87.
Manjikian M. Cybersecurity ethics. Routledge; 2018.
Martin CD. White hat, black hat. ACM Inroads 2017;8(1):33–5.
McNamara A, et al. Does ACM's code of ethics change ethical decision making in software development?. Proceedings of the ESEC/FSE 2018, 2018.
Meyer P. Norms of responsible state behaviour in cyberspace. In: The Ethics of Cybersecurity. Springer; 2020. p. 347–60.
Mirkovic J, Reiher P. A taxonomy of DDoS attack and DDoS defense mechanisms. ACM SIGCOMM Comput. Commun. Rev. 2004;34(2):39–54.
Molitorisz S. Net Privacy. NewSouth Publishing; 2020.
Morgan G, Gordijn B. A care-based stakeholder approach to ethics of cybersecurity in busines. In: The Ethics of Cybersecurity. Springer; 2020. p. 119–38.
Mouton F, et al. Necessity for ethics in social engineering research. Comput. Secur. 2015;55:114–27.
Newell BC. The massive metadata machine. J. Law Policy Inf. Soc. 2014;10(2):481–522.
Nissenbaum H. Where computer security meets national security. Ethics Inf. Technol. 2005;7(2):61–73.
Nissenbaum H. A contextual approach to privacy online. Daedalus 2011;140(4):32–48.
Opie R. Gifted individual. ABC News; 2019 https://www.abc.net.au/news/2019-05-27/adelaide-teenager-hacked-into-apple-twice-in-two-years/11152492.
Penney J. Chilling Effects. Berkeley Technol. Law J. 2016;31(1):117–61.
Prosser WL. The right to privacy. Calif. Law Rev. 1960;48:383–423.
Quarterman JS. System Administration As a Criminal activity, Or the Strange Case of Randal Schwartz, 5. Matric News; 1995 https://groups.csail.mit.edu/mac/classes/6.805/articles/computer-crime/schwartz-matrix-news.txt.
Rawls J. A Theory of Justice. Harvard University Press; 1971.
Rest JR, et al. Postconventional Moral Thinking. Lawrence Erlbaum; 1999.
Roberts ST. Digital detritus: "Error" and the Logic of Opacity in Social Media Content moderation. First Monday; 2018. doi:10.5210/fm.v23i3.8283.
Richards D, et al. A proposed AI-enhanced serious game for cybersecurity ethics training. Proceedings of the 9th Conference of the Australasian Institute of Computer Ethics, 2020. https://auscomputerethics.files.wordpress.com/2021/03/aice_2020_paper_1-1.pdf.
Scarfone KA, et al. Technical Guide to Information Security Testing and assessment. National Institute of Standards and Technology; 2008. doi/10.6028/NIST.SP.800-115.
Schlehahn E. Cybersecurity and the state. In: The Ethics of Cybersecurity. Springer; 2020. p. 205–26.
Shanley A, Johnstone M. In: 13th Australian Information Security Management Conference. Selection of penetration testing methodologies Perth; 2015.
Shea M. Forty years of the four principles. J. Med. Philos. 2020a;45(4–5):387–95.
Shea M. Principlism's balancing act. J. Med. Philos. 2020b;45(4–5):441–70.
Shires J. Enacting expertise. Politics Gov. 2018;6(2):31–40.
Simone MA. Give me liberty and give me surveillance. Crit. Discourse Stud. 2009;6(1):1–14.

Solove DJ. Understanding Privacy. Harvard University Press; 2008.
Staines D, Formosa P, Ryan M. Morality play. Games Culture 2019;14(4):410–29.
Stevens S. A framework for ethical cyber-defence for companies. In: The Ethics of Cybersecurity. Springer; 2020. p. 317–29.
Tavani H. The conceptual and moral landscape of computer security. In: In *Internet Security*. Jones & Bartlett Learning; 2007. p. 29–45.
Taylor J, Wong J. Cloudflare Cuts Off Far-Right Message Board 8chan After El Paso Shooting. The Guardian; 2019 https://www.theguardian.com/us-news/2019/aug/05/cloudflare-8chan-matthew-prince-terminate-service-cuts-off-far-right-message-board-el-paso-shooting.
Thomson JJ. The right to privacy. Philos. Public Aff. 1975;4(4):295–314.
Timmers P. Ethics of AI and cybersecurity when sovereignty is at stake. Minds Mach. 2019;29(4):635–45.
Vallor S. An Introduction to Cybersecurity Ethics. Markkula Center for Applied Ethics; 2018 https://www.scu.edu/media/ethics-center/technology-ethics/IntroToCybersecurityEthics.pdf.
Van de Poel I. Core values and value conflicts in cybersecurity. In: The Ethics of Cybersecurity. Springer; 2020. p. 45–72.
Vaniea K, Rashidi Y. Tales of Software Updates. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016.
von Ahn L, et al. CAPTCHA. In: Biham E, editor. In: Advances in cryptology—EUROCRYPT. Springer; 2003. p. 294–311.
Warren SD, Brandeis L. The right to privacy. Harvard Law Rev. 1890;4/5:193–220.
Wash R, et al. In: Symposium On Usable Privacy and Security (SOUPS). Out of the Loop. Menlo Park; 2014.
Weber K, Kleine N. Cybersecurity in health care. In: The Ethics of Cybersecurity. Springer; 2020. p. 139–56.
Wilk A. In: IEEE International Conference on Software Science, Technology and Engineering. Cyber security education and law Beer- Sheva; 2016.
Weizenbaum J. On the impact of the computer on society. Science 1972;176(4035):609–14.
Zajko M. Security against surveillance. Surveill. Soc. 2018;16(1):39–52.

1. Paul Formosa, Department of Philosophy, Macquarie University

Paul Formosa is an Associate Professor in the Department of Philosophy at Macquarie University, and the Director of the Centre for Agency, Values and Ethics. Paul has published widely in topics in moral and political philosophy with a focus on Kantian ethics, the nature of evil, and the ethical issues raised by videogames, technology, cybersecurity and AI. His-work has been published with Oxford and Cambridge University Presses and in journals such as Ethics and Information Technology, Games and Culture, European Journal of Philosophy, and Ethical Theory and Moral Practice.

2. Michael Wilson, School of Law, Murdoch University

Michael Wilson is a Lecturer in the School of Law at Murdoch University. He was awarded his PhD in 2020 for a thesis examining the problem of 'going dark' and Australian digital surveillance law. His research interests include the regulation of cryptography, computer hacking, cybersecurity ethics, surveillance law, and digital evidence. He has published in journals such as Crime, Law and Social Change, International Communication Gazette, and Trends & Issues in Crime and Criminal Justice.

3. Deborah Richards, Department of Computing, Macquarie University

Deborah Richards is a Professor in the Department of Computing at Macquarie University. Following 20 years in the IT industry during which she completed a BBus (Comp and MIS) and MAppSc (InfoStudies), she completed a PhD in artificial intelligence on the reuse of knowledge at the University of New South Wales and joined academia in 1999. While she continues to work on solutions to assist ethical decision-making and knowledge acquisition, for the past decade, her focus has been on intelligent virtual agents, virtual worlds and serious games for education, health learning and well-being to challenge attitudes and empower users to make good choices.