Ethical Frameworks and Computer Security Trolley Problems: Foundations for Conversations

Tadayoshi Kohno, *University of Washington;* Yasemin Acar, *Paderborn University & George Washington University;* Wulf Loh, *Universität Tübingen*

https://www.usenix.org/conference/usenixsecurity23/presentation/kohno

This paper is included in the Proceedings of the
32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

Ethical Frameworks and Computer Security Trolley Problems: Foundations for Conversations

Tadayoshi Kohno
University of Washington

Yasemin Acar
Paderborn University &
George Washington University

Wulf Loh
Universität Tübingen

Abstract

The computer security research community regularly tackles ethical questions. The field of ethics / moral philosophy has for centuries considered what it means to be "morally good" or at least "morally allowed / acceptable". Among philosophy's contributions are (1) frameworks for evaluating the morality of actions — including the well-established consequentialist and deontological frameworks — and (2) scenarios (like trolley problems) featuring moral dilemmas that can facilitate discussion about and intellectual inquiry into different perspectives on moral reasoning and decision-making. In a classic trolley problem, consequentialist and deontological analyses may render *different* outcomes. In this research, we explicitly make and explore connections between moral questions in computer security research and ethics / moral philosophy through the creation and analysis of trolley problem-like computer security-themed moral dilemmas and, in doing so, we seek to contribute to conversations among security researchers about the morality of security research-related decisions. We explicitly do *not* seek to define *what* is morally right or wrong, nor do we argue for one framework over another. Indeed, the consequentialist and deontological frameworks that we center, in addition to coming to different conclusions for our scenarios, have significant limitations. Instead, by offering our scenarios and by comparing two different approaches to ethics, we strive to contribute to *how* the computer security research field considers and converses about ethical questions, especially when there are different perspectives on what is morally right or acceptable. Our vision is for this work to be broadly useful to the computer security community, including to researchers as they embark on (or choose not to embark on), conduct, and write about their research, to program committees as they evaluate submissions, and to educators as they teach about computer security and ethics.

1 Introduction

We believe in the essentiality of maintaining high ethical standards when conducting and evaluating computer security research. As examples of the field's[1] commitment to moral considerations, conference program committees are leveraging ethics review boards and authors are discussing ethics in submissions. There also exist tools to help community members make adequate moral decisions, such as the 2012 Menlo Report [56] and recent author guidelines in security conference calls for papers.

However, challenges still arise. Central to these challenges is that in some cases there may *not* be universal agreement on what constitutes an adequate decision. Consider, for example, a hypothetical scenario (our Scenario A) in which researchers find a vulnerability in a wireless implantable medical device. Assume that the device manufacturer is out of business and, hence, it is impossible to patch the vulnerability. Also, assume that there is *zero* chance of the vulnerability ever being exploited, even if adversaries know about it.[2] Should the researchers *disclose the vulnerability* to the government and the public, thereby respecting patients' right to be informed (a key component of the "respect for persons" principle of the Menlo Report [56] as well as the earlier Belmont Report [54] and the principle of "autonomy" in the Principles of Biomedical Ethics [7])? Or, should the researchers, knowing that adversaries would *never* manifest but that a knowledge of the vulnerability's existence could harm patients (who might remove the device from their bodies and hence lose the health benefits out of unnecessary concerns), *not disclose* the vulnerability to the government and the public (thereby respecting the principle of "beneficence" and the avoidance of harm, another core element of the Menlo Report [56], the Belmont Report [54], and the Principles of Biomedical Ethics [7])?

There are strong arguments for *both* decisions. In a situation with conflicting arguments, how are we as a field to make

[1] When we say "the field", we refer to the computer security research field even though our team is composed of both computer security researchers and a moral philosopher. We use this terminology because our primary goal is to contribute to the computer security research field.

[2] To enable us to focus on the philosophical aspects of ethics and morality and not become entangled in real-world details, we make simplifying assumptions in this and all our scenarios. We elaborate on this decision in Section 3.

the right decision? We argue that whatever process is used, that process will benefit from being informed by philosophy's understanding of the different approaches that people take to ethics — approaches that *can* result in different people coming to different conclusions. Hence, our research. In short, we seek to contribute to future conversations about *what* is morally right, good, or allowed, and we do so by studying *how*, from a philosophical perspective, to have such discussions.

Ethics and Moral Philosophy. The field of ethics / moral philosophy centers the question of what it means to be "good", "morally right", or "morally allowed" (i.e., not prescribed but also not forbidden). Even in the field of philosophy, there is no consensus for *what* this exactly means. But, there is a mature understanding for *how* to discuss what it might mean. Conversations about "good" and "bad" begin by centering a perspective — an *ethical framework*.

Ethical Frameworks. Ethical frameworks define approaches for reasoning about what is morally right or wrong, i.e., what is "good" or "bad". Two of today's leading frameworks (or, more precisely, categories of frameworks) are *consequentialist* and *deontological ethics*. Consequentialist ethics centers questions about the impacts (consequences) of different decisions. Under consequentialist ethics, one might assess the benefits and harms of different options before making a decision that maximizes net benefits. Deontological ethics centers questions about duties (deon) and rights. Under deontological ethics, one might ask what rights different stakeholders have, e.g., a right to privacy or a right to autonomy.

We center consequentialist and deontological ethics — and in particular utilitarianism and Kantian deontological ethics, respectively (Section 4) — in our study because of (1) their prominence in the field of ethics / moral philosophy (they are two of the three leading frameworks) and (2) their existing impact on the computer security research field's approach to ethics and morality (e.g., the Menlo Report [56] derives from the Belmont Report [54], which itself embeds both consequentialist and deontological elements). We stress, however, that both consequentialist and deontological ethics have limitations and that by centering them we are *not* arguing that anyone adopt a strict consequentialist or deontological perspective. At a minimum, one might include considerations from both frameworks, as the Menlo Report [56] does. As we discuss more in Section 3, modern frameworks include a more critical perspective. Additionally, much of philosophy's discussion of consequentialist and deontological ethics centers a Western perspective. While Western frameworks encompass ethical considerations that are part of non-Western traditions (e.g., about duties towards each other, the nature of fundamentally relating to each other, the outcomes of actions / policies, and so on), each tradition has its own unique history and elements. Although outside the scope of this work, we encourage the computer security research community to gain greater familiarity with other frameworks as well.

Our Work. As exemplified by the Menlo Report [56] and recent calls for papers, there are *already* connections between the computer security research field and ethics / moral philosophy. Our assessment is that many of these connections are implicit. We seek to make these connections explicit and, by doing so, contribute to *how* the field discusses and considers moral questions.

To do our research, we composed a team of researchers consisting of both those trained in computer security research *and* those trained in moral philosophy. The computer security researchers on our team have significant prior work addressing and discussing ethical questions in computer security research. However, prior to this collaboration, the security researchers approached ethics from a "we should be good" and "having thought carefully and talked with others, I think this is right" approach rather than from an approach informed by ethics / moral philosophy. The moral philosopher on our team has significant prior work in applied ethics outside of computer security. Our work is thus cross-disciplinary and could be read as both a work in philosophy (particularly normative and applied ethics) and (we believe) a contribution to the computer security research field.

Goals, Methods, and Findings. We seek to leverage tools and insights from ethics / moral philosophy to facilitate clear, thoughtful, and rigorous conversations within the computer security research field about what are morally right or allowed decisions / policies / institutions — i.e., we seek to contribute to *how* the field discusses moral questions. We do *not* seek to define *what* (morally) "right" or "good" means (which would be a metaethical question).

Our methodology is to:

- Develop computer security scenarios reminiscent of classical ethical dilemmas and for which evaluations under different ethical frameworks justify different outcomes; our scenarios are akin to philosophy's classic trolley problems, which we describe later. (Section 3.)

- Explore those computer security scenarios using both consequentialist (utilitarianism) and (Kantian) deontological analyses. (Section 5.)

- Develop additional computer security scenarios that, individually, may not pose ethical dilemmas (i.e., there may be stronger agreement for what constitutes a morally right or allowed decision for some scenarios) but that, together, facilitate deeper explorations about moral considerations within the field. (See the full version of this paper [34].)

- Reflect upon the above scenarios and explorations and derive lessons about *how* to have informed conversations about ethics and morality in the computer security research community. (Section 6.)

Summary of Our Three Main Scenarios. We summarized Scenario A above. Scenario B explores the morality of studying stolen data — data that people did not intend to be public. Scenario C explores what to do if a program committee member encounters a submission containing undisclosed information about their company's product. All our scenarios are based on actual situations encountered within the computer security research community, though we modified the scenarios to make them more conducive to ethical analyses, per our research goals. While reflective of real-world scenarios, our scenarios do not cover the full spectrum of moral dilemmas encountered within the security research community, nor is it our intent to do so.

Example Use Case of Our Results: Program Committee Discussions. The security researchers on this team have on multiple occasions encountered the following situation:

- A paper is submitted to a peer-reviewed conference. The paper reports on work that one program committee member flags as possibly unethical.
- Program committee members discuss the morality of the work but cannot agree; some committee members think it was ethical, and others think it was not.

Such disagreements can be challenging if, for example, some committee members adopt consequentialist perspectives and other committee members adopt deontological perspectives, but the committee members do not realize that they are using different frameworks for evaluating morality. Prior to this collaboration, we (the security researchers on this team) did not have the tools and language to untangle such disagreements. Now, through this collaboration and the exploration of computer security scenarios via the consequentialist and deontological frameworks, we have that language.

Example Use Case of Our Results: Discussing Research Path. In many cases, there may already exist clarity for researchers on how to navigate moral questions, e.g., researchers might follow the recommendations in the Menlo Report [56]. However, there may remain times when clarity does not exist, e.g., when there are tensions between what is morally right from a benefits / harms perspective (consequentialist ethics) and what is right from a duties / rights perspective (deontological ethics). Through the articulation of established ethical frameworks, and through the exploration of computer security scenarios via these frameworks, we hope to help researchers have more methodical and informed discussions about ethics and morality when there are such tensions. Since different frameworks can lead to different conclusions of what is morally right, however, we stress that the frameworks should *not* be used to justify a path that researchers have *a priori* decided that they want to take. Rather, we argue that the ethically correct process is to *center* ethics in the decision of whether or not to do a research project or do some component of the research and accept that sometimes the answer is "no".

Publication Information. The full version of this paper is available online at https://securityethics.cs.washington.edu/ [34].

2 Motivation and Background

2.1 Ethics / Moral Philosophy

Ethics / moral philosophy is a field that has existed for centuries. In Western culture, the most well-known ethical frameworks are virtue ethics (most notably developed by ancient Greek philosophers such as Plato and Aristotle), deontological ethics (a famous example from German Enlightenment philosopher Immanuel Kant), and utilitarianism (an example of consequentialist ethics, first developed by Jeremy Bentham and John Stuart Mill). In other cultures, classic ethical frameworks include Confucianism, Daoism (as first coined by Laozi), and Ubuntu.

As a field that is centuries old and spans cultures and histories, it is natural that there is no universal consensus on what, precisely, ethics and morality mean. For our work, we use *ethics / moral philosophy* to refer to the (scientific) exploration of *how* to consider, evaluate, and discuss moral questions,[3] and *morality* to refer to the object of this exploration. For example, ethicists / moral philosophers use ethics in the sense of moral reasoning to determine whether an action, social institution, or set of norms is moral or not [22].

Modern ethicists often use *ethics* and *morality* interchangeably.[4] Thus, the computer security field is not wrong in its use of the term *ethics* to encompass both ethics and morality. When precision is not necessary, we may do so as well.

We provide a deeper background on ethical frameworks in Section 4.

2.2 Ethics and Computer Security Research

The field of computer security research has a long history with questions of ethics and morality. This history includes research directions that required significant ethical forethought and planning before implementation as well as research projects that, upon completion, raised concerns within the community. We touch on elements of this history in more detail in the full version of this paper [34]. We focus here on historical elements particularly relevant to our philosophical explorations.

In the U.S., the 1976 Belmont Report [54] serves as a foundation for the ethical treatment of research subjects (human subjects) within universities. European universities and other research institutions may reference general "good scientific

[3] Whereas "moral philosophy" refers to a mainly philosophical endeavor, "ethics" also comprises non-strictly-philosophical (e.g., theological) reasoning.

[4] Against this, Habermas argues that "ethics" refers to questions about the good life, whereas "morality" is concerned with what we owe others [25].

A Classic Dilemma: The Trolley Problem

Context:

- A runaway trolley with no brakes is heading straight along a set of tracks.
- Five people are tied to those tracks.
- One person is tied to an alternate set of tracks.
- A trolley operator has the ability to change the trolley's path and make it head down the alternate set of tracks.

The choice for the trolley operator:

- *Do nothing:* Five people die.
- *Make the trolley take the alternate set of tracks:* One person dies.

Figure 1: The trolley problem is a classic thought experiment / ethical dilemma.

practices", and may discuss dual use of research results (civilian and military), but often omit formal ethics reviews of research ideas, leaving the ethical decisions to researchers.

The 2012 Menlo Report [56] applies the Belmont Report's principles of justice, beneficence, and respect for humans (not just human subjects) to computer security research. In doing so, the Menlo Report highlights that computer security research may impact computer systems and their users beyond consenting participants. As an applied ethics framework, the Menlo Report [56] makes ethical practices and thought accessible to a broad audience. The Menlo Report [56] is now explicitly referenced in the 2022 and 2023 IEEE Symposium on Security & Privacy and the 2023 USENIX Security call for papers. Our work is motivated, in part, by our belief that the application of the Menlo Report, and security ethics conversations in general, can be further enriched with a greater understanding of ethics / moral philosophy.

2.3 A Classic Moral Dilemma

Ethicists / moral philosophers have, for generations, proposed dilemmas for ethical debate and consideration. A classic dilemma (or, more precisely, family of dilemmas) are the "trolley problems". These are dilemmas because they present a choice between two options, both of which contain undesired aspects. Therefore, different ethical frameworks potentially present different answers to such dilemmas. Some authors (among them Philippa Foot herself, who came up with the original trolley problem [20]) take them to show that people's moral intuitions will most likely diverge in important cases.

Figure 1 presents an archetypical trolley problem. In this trolley problem, a runaway trolley with no brakes is heading straight down a track. Five people are tied to that track. A trolley operator is watching the trolley. They could do nothing, in which case five people would die. The trolley operator could, however, choose to redirect the trolley down a second, adjacent track. If the operator does so, then the trolley would kill only one person — the person tied to that adjacent track.

Philosophers and psychologists have studied people's responses to trolley problems such as in Figure 1 and, indeed, there is no universal consensus for what constitutes the morally correct action of the trolley operator [20]. In psychology studies, for example, differences can arise due to the moral intuitions and values of the participant and may vary by culture, e.g., [3, 12, 13, 24, 36, 59].

Variants of the trolley problem feature different outcomes and can elicit different thought processes and decisions.[5] As an example variant, the single person on the alternate track might be a young child whereas the five people on the main track might already be near death. As another variant, the five people tied to the main track might have tied themselves there intentionally whereas the single person on the other track might be there against their will. Or, the five people on the main track might have been convicted of war crimes by an international tribunal whereas the person on the alternate track is known to have led a virtuous life.

3 Computer Security Trolley Problems

3.1 Scenario Generation Process

Our research team used a collaborative and interactive process for scenario generation. After discussing our initial approach, we present our final methodology and scenario selection criteria.

Initial Approach. Initially, the security researchers on the team created scenarios representative of scenarios that we (the security researchers) had previously encountered (e.g., as program committee members or as researchers). Our team

[5] An interactive exploration of different trolley problems is available at https://neal.fun/absurd-trolley-problems/.

generated dozens of such scenarios with an initial goal of exhaustively and systematically surfacing the full spectrum of ethical considerations encountered within our field. While the generative process was important toward normalizing an understanding of scenarios and moral issues in the field across the entire team (including both the security researchers and the philosopher), these scenarios had several key limitations:

- **Too late.** Some scenarios were framed as "a program committee member reads a conference submission in which the authors did such-and-such; the program committee member believes that such-and-such should not have been done; should the program committee accept the paper?"
- **Open-ended.** Other scenarios were very open-ended and highly unconstrained, e.g., "here is an issue that a research group encountered, what should they do?"
- **Not a dilemma.** Some scenarios had relatively clear and uncontroversial moral implications; we encountered them as program committee members because (for example) of an oversight by the authors of a paper submission, e.g., because the authors assumed that the IRB process was sufficient to cover all aspects of moral decision-making.
- **Indecisive.** Some of our scenarios did not have conclusive decisions under different ethical frameworks, at least not without significant additional information that would greatly expand the scenarios and make them unwieldy.

The "too late" scenarios all shared a common theme: researchers made decision X, for some X; what should the program committee do if they question the morality of X? A discussion of the ethical processes for program committees when encountering such papers is important, and indeed we consider a family of such scenarios in the full version of this paper [34]. For our core ethical dilemmas (this section), we sought scenarios featuring a decision *before* a controversial act X is committed in the first place. As a concrete example, for our Scenario B (to be described), an initial version featured a scenario in which a program committee reviews a paper that studies data that some program committee members believe should not have been studied. What should the program committee do? Our final Scenario B asks: should researchers study that data?

The "open-ended" scenarios, while representative of what researchers might encounter in the real world, made analyses of the scenarios under established ethical frameworks too unconstrained for focused treatments. As evidenced by our team's internal discussions, when faced with open-ended questions of the form "what should the researchers do?", it is possible to spend hours, and hence volumes of written pages, exploring different possible paths forward. While for some scenarios such explorations would be important contributions of their own, those are not the contributions we sought with this work. Rather, we wanted scenarios conducive to short, precise, and focused analyses with minimal (binary) options.

The "not a dilemma" scenarios were intellectually interesting and important in establishing our team's shared understanding of questions of morality and computer security research. However, because these scenarios were not actual dilemmas, evaluation under different ethical frameworks resulted in the same conclusions and hence were not as generative of philosophical explorations as scenarios that yielded different conclusions under different ethical frameworks. In short, we sought scenarios for which people — including computer security research community members — might, through sound reasoning, plausibly disagree.

The "indecisive" scenarios featured decisions for which all possible choices would result in "comparable" benefits / harms that would need extensive empirical work to assess. In the real world, if one were to encounter such a situation, a significant portion of the conversation might center on assessing those empirical claims. For our work, we wanted to center ethical and moral thought processes, not empirical questions. Hence, we sought scenarios without complicated benefits / harms calculus.

Revised Approach: Criteria, Creation, and Validation. Informed by the results of our analyses of and conversations about our initial scenarios, our team developed the following criteria for scenario generation:

- **Early.** We sought scenarios that featured moral questions that actors (e.g., researchers) might encounter about their own future actions, not questions about what to do after it has been determined that researchers have already committed a morally questionable act.
- **Binary options.** We sought scenarios that — like the trolley problems — have binary options for some actor (e.g., the trolley operator in the trolley problem in Figure 1 or a research team in computer security-related scenarios).
- **Dilemmas.** We sought scenarios that were true dilemmas. Specifically, we sought scenarios for which analyses under consequentialist and deontological ethics would yield different conclusions.
- **Decisive.** We sought scenarios for which analyses under the consequentialist and deontological ethical frameworks were clear, straightforward, and decisive. Sometimes this came at the cost of simplifying and artificially contrasting the ethical traditions to bring out key differences in perspective and focus.

Our research team iterated extensively on the creation of scenarios that satisfied these criteria, over regular meetings throughout late summer and fall 2022 and early 2023. Our iteration was both at a high level, focusing on the scenario's overall setup and context, and at a low level, focusing on fine nuances and details. As we iterated on these scenarios, we presented variants in university seminars (at other universities) and in courses (at the undergraduate and graduate levels). After each presentation, we reflected upon and revised the scenarios as needed to address ambiguities or clarify key aspects relevant to the scenarios' intended moral questions. We additionally shared our scenarios with others in the computer security research community for feedback.

In addition to being instrumental to the process of scenario creation, this iterative process also served as scenario validation. Specifically, the iterative process with systematic philosophical analyses and external discussions helped us validate that our scenarios met our "dilemmas" and "decisive" criteria. (That our scenarios met the "early" and "binary options" criteria was easy to assess by construction.)

Based on Reality, But Not Real. We stress that although our final scenarios are based on reality, they are *not* realistic. Real-world scenarios generally do not present only a binary option to decision-makers — they present a medley of options. Additionally, to enable precise analyses under different ethical frameworks, our scenarios minimize uncertainty. The real world, on the other hand, is full of uncertainty, e.g., uncertainty about when or if an adversary might manifest or the actual benefits / harms of a technology or exploit. Thus, assessing benefits / harms (for consequentialist ethics) and rights violations (for deontological ethics) is significantly more challenging in the real world than in our scenarios. Real-world scenarios may have multiple actors simultaneously making decisions, each of which might impact the other actors; in our scenarios, we consider only a single decision-maker. Additionally, to simplify our analyses, we reduce the impacts of decisions on the decision maker in our core scenarios (Scenarios A, B, and C); we add such impacts into some of the supplementary scenarios in the full version of this paper [34]. In the real world, decision-makers may involve others in the decision-making process; our scenario descriptions do not preclude such discussions but leave the final decision in the hands of the specified decision-maker rather than allow for the transference of the decision responsibility to another entity (e.g., a committee or government).

The Structure of a Scenario. For each scenario, we use a structure similar to Figure 1 for the trolley problem. Each scenario centers a decision-maker and has:

- **Context:** The "context" of the scenario provides the background context for the decision that the actor needs to make.
- **Choice:** The "choice" of the scenario describes two options that the actor must choose between.

We use prose to describe the context and choice. The full version of this paper [34] provides figures, like Figure 1, for each scenario.

3.2 Scenario A: Medical Device Vulnerability

Scenario A centers around researchers who discover a vulnerability in a wireless implantable medical device.

Context. Researchers found a vulnerability in a wireless implantable medical device made by a manufacturer that is no longer in business. Existing patients still use the device and new patients are still receiving the device. It is not possible to update the software on the device and patch the vulnerability. Even if the researchers disclose the vulnerability to the public, there is zero probability of the vulnerability being exploited in the wild. There are no field- or industry-wide gains to be made via the public disclosure and discussion of the vulnerability, e.g., the public disclosure of the vulnerability would not teach the field any new lessons about computer security and medical devices.

The Choice. For this scenario, a disclosure to some sufficiently large group (e.g., all healthcare professionals who work with the relevant medical condition) would eventually result in a disclosure to the public (through information leakage). Hence, the researchers must choose between not disclosing the vulnerability to anyone or disclosing the vulnerability to the government, the healthcare industry, and the public.

If the researchers disclose the vulnerability to the public, then patients may be harmed psychologically (a fear of having a vulnerable / imperfect device even if the likelihood of it being compromised is zero) or physically (the device increases a person's life by ten years; if a patient removes or does not receive the device, they would not receive the health benefits).

If the researchers do not disclose the vulnerability to anyone, then patients do not have the option to make an informed choice with respect to whether they keep the device or, for new patients, whether or not they receive the device.

On this Scenario. In 2008, one of us (T.K.) co-authored a study that discovered and reported on vulnerabilities in a wireless implantable medical device [26]. We thought deeply about ethics and responsible disclosure at the time of that study, and the medical device security field has continued to reflect upon ethics and responsible disclosure thereafter, e.g., [35, 46]. We designed Scenario A to center patient-focused elements of consideration: the fundamental rights that patients have and the benefits and harms to patients with either disclosing or not disclosing a vulnerability. To center the ethical considerations on the patients, in Scenario A it is not possible to update the software on the medical device, and hence a traditional coordinated disclosure process of first notifying the manufacturer and then giving them time to respond is not an option (a situation which, unfortunately, is plausible [49]). Additionally, the healthcare industry has already internalized the importance of computer security for wireless implantable medical devices, e.g., [15, 55], and hence there are no significant field-wide positive impacts with a public disclosure. To meet our scenario design criteria, this scenario presents only two options to the researchers. In a real-world scenario, we anticipate much greater involvement from organizations like the U.S. Food and Drug Association (FDA); the researchers might even cede the final decision to the FDA or another entity such as U.S. CERT. Additionally, factors such as FDA policy might impact the plausibility of Scenario A.

3.3 Scenario B: Studying Stolen Data

Scenario B centers researchers who are trying to decide whether or not to study stolen data.

Context. Company B offers a service that matches job applicants with jobs. The public believes that Company B's AI matching system has racial and gender biases. Some people also believe that Company B's AI system could be manipulated by adversaries. Adversaries compromise Company B's servers and steal the entirety of their data, including all data about all past job postings, all past job application packets, and the outputs of all past job-applicant matches from Company B's AI system. The adversaries also steal all internal details of Company B's AI matching system, including the underlying ML model. The thieves post the stolen material online; a research group obtains a copy of the stolen material as soon as it is publicly available. Subsequently, many victims of the data breach — the job applicants — publicly state their desire for the stolen data to be permanently deleted, everywhere; all publicly-available copies are then deleted.

The Choice. The research group wishes to study the stolen data and scientifically assess whether Company B's AI matching system is, in fact, biased. If it is biased, the researchers seek to measure past impacts of those biases, e.g., by counting the number of applicants not forwarded to employers because of racial or gender biases. Additionally, using the stolen data — including both the ML model and knowledge of the contents of past application packets — the researchers hope to assess the vulnerability of Company B's AI system to adversarial manipulation. Informed by a scientific understanding of the biases and vulnerabilities in Company B's AI system, the researchers intend to propose technical and policy mechanisms to mitigate such biases and vulnerabilities in the future.

The researchers know, however, that the data was stolen and shared publicly over the objections of many job applicants. The researchers must choose between doing nothing (not studying the data) or studying the data and reporting on the results. If the researchers study the data and report on their results, they know not to include anything in their publication that could lead to the identification of any of the job applicants. If they study the data, they also know that they must continue to retain a copy of the data even after publishing their results in case their results are challenged, e.g., by Company B.

On this Scenario. Adjacent to the computer security research field, the human-computer interaction field has an extensive history of considering the morality of studying data that people might have technically made public but that they might not wish to be used in research or that might cause harms if quoted in a publication, e.g., [9, 17, 18, 39, 43, 60]. These works also consider best practices for how to study such data and how to report on the results.

Within the security research community, it is not uncommon to study datasets containing information that users did not intend to be public. A typical example is the study of the contents of stolen password or other databases [52]. An adjacent example is the study of anonymized datasets that are, in actuality, not fully anonymized, e.g., [2, 40, 51]. The ubiquity of such studies speaks to at least partial agreement within the community on the morality of such studies in general, though researchers must still pay attention to details. For example, even if researchers study the contents of a leaked password database, they might not include real username and password pairs in a resulting publications, similar to how human-computer interaction researchers might not include full quotes in publications even if quoting from public data, e.g., [5, 9, 18, 39].

For Scenario B, we sought a scenario related to stolen data but with content that, by itself, is more sensitive than usernames and passwords. Motivated in part by past computer security research on biases and vulnerabilities in remote proctoring software [10] as well past concerns about biases in job-applicant matching systems, e.g., [58], we chose to focus on an AI job-applicant matching system: a system for which job applicants might submit an extensive amount of private information.

As with all our scenarios, our goals in Section 3.1 influenced our scenario design. Here we highlight two aspects of this scenario that enable it to meet our "decisive" goal. First, while one might argue that people's right to privacy extends to data that they intended to be private even after others (illegally) made the data public, we make the right to privacy in Scenario B even more definitive by having those impacted by the data leak explicitly request that all copies of the data be deleted. Second, if biases are present in the AI system, and if those biases are removed, that would change which applicants are shown to employers. Although preferable in terms of overall fairness, such a change could also do harm, e.g., to the removed applicants. To simplify our consequentialist analyses, we explicitly assume that anyone removed through this process would still be able to find a job that they desire.

3.4 Scenario C: Inadvertent "Disclosure"

Scenario C features an ethical dilemma for a conference program committee member. We selected this scenario to be among the three featured because questions of ethics and morality arise not only in research (Scenarios A and B), but also during the peer review process (this scenario).

Context. A program committee member works for Company C and, as part of the program committee process, encounters a confidential paper submission detailing an undisclosed vulnerability in Company C's product. Upon reading the submission, the Company C employee realizes that the vulnerability is very serious and that it will take a significant amount of time to patch. The employee feels an obligation to their employer and to Company C's users. But, the program chairs required all committee members to explicitly agree to maintain the confidentiality of all submissions. Company C's leadership team decided that the Company C employee should agree to the confidentiality condition and join the program committee.

The Choice. The employee of Company C must decide between doing nothing (not disclosing the vulnerability in the paper to Company C) or disclosing the vulnerability to their employer.

On this Scenario. While we are aware of real-world scenarios similar to Scenario C, we are unaware of written public statements about those situations and consequently include no background citations. Scenario C is thus based solely on the memories and experiences of the computer security researchers on this team as well as discussions with others. As with our other scenarios, the real world is more complex, with additional options available to the program committee member, e.g., the program committee member could work with the program chairs to determine a course of action.

4 Ethical Frameworks

Ethical frameworks define approaches for reasoning about whether actions are morally right or wrong. In ethics / moral philosophy, the oft-cited three main ethical frameworks are consequentialist ethics, deontological ethics, and virtue ethics. A fourth oft-discussed framework is discourse ethics. We discuss the first two in Section 4.1. Although our analyses focus on the first two, we discuss the latter two along with several other frameworks, including principlism (featured in the Belmont and Menlo Reports [54, 56]) in Section 4.2.

The frameworks we explore have in some cases evolved over considerable periods of time, with a multitude of contributions, objections, and adaptations. There can thus exist a vast variety of different branches and nuances within each framework. Since our goal is to explore moral dilemmas in computer security research from the perspective of different ethical frameworks and not to argue, for example, the benefits of one framework over another or for a new theory of ethics for computer security research, we limit our descriptions to the general features of each framework. Our summaries are sufficient to clearly contrast the different frameworks with each other and to receive clearly distinguishable reasonings and outcomes with regard to our scenarios from Section 3.

While we believe that our summaries are sufficient to enable security researchers to explore their own problems with these frameworks, we defer interested readers to works such as Anscombe's article "Modern Moral Philosophy" [4], Baggini and Fosl's book *The Ethics Toolkit* [6], Deigh's book *An Introduction to Ethics* [14], Driver's book *Ethics: The Fundamentals* [16], and Stanford University's online resources [48] for additional, general information. For works focused on ethics and technology / engineering, we defer readers to works such as Floridi's book *The Cambridge Handbook of Information and Computer Ethics* [19], Iphofen's book *Handbook of Research Ethics and Scientific Integrity* [32], Quinn's book *Ethics for the Information Age* [45], and Santa Clara University's online resources [57], as well as professional codes of ethics [1, 31, 41]. Further, as we stress elsewhere, we are *not* arguing for the application of any of the frameworks we survey; rather, we are arguing for the use of these frameworks as mechanisms to facilitate thoughtful dialog and inquiry while, for example, applying the principles in the Menlo Report [56].

4.1 Consequentialist and Deontological Ethics

As discussed earlier, we center consequentialist and deontological ethics in our analyses because of their prominence in the field of ethics / moral philosophy and because of their existing role in the computer security research community, e.g., their presence in the Menlo Report [56].

Consequentialist Ethics. Consequentialism centers the consequences — the outcomes — of an action, both positive (benefits) and negative (harms). Each consequentialism comprises of a value theory (e.g., hedonism) and a moral principle (e.g., maximizing overall utility), according to which an action is morally right exactly when there is no other action with better consequences as measured by the respective value theory.

Utilitarianism is an example of consequentialism in which positive and negative outcomes are generally assessed with respect to the well-being (welfare) of people. We use utilitarianism in the consequentialist analyses in this paper. Under utilitarianism, the right action is the action that produces the greatest net positive well-being. There are three main categories of utilitarianism,[6] each corresponding to one of three main theories of well-being:

- **Hedonic utilitarianism:** An action is right if it produces the greatest net happiness — the greatest aggregate happiness over a given set of individuals [50].
- **Preference utilitarianism:** An action is right if it enables the greatest number of people to live by their own preferences [27].
- **Objective list utilitarianism:** An action is right if it produces the greatest net positive impacts on the greatest number of people with respect to an objective list of measures [42]; example measures are the levels of one's health, wealth, or access to resources.

[6] For the purposes of this paper and the decisions in the scenarios, we focus on direct action utilitarianism and ignore rule utilitarianism.

For objective list utilitarianism, the standards to maximize for are not subjective desires or preferences, but rather "objective" (in the sense of "applicable to all") measures such as level of health, wealth, and safety (happiness could also be one standard on the list, though some argue that happiness cannot be objectively measured).

These categories are related but distinct. For example, increased health (an objective list measure) can lead to increased happiness (the hedonic measure). Likewise, if someone can live by their preference (preference), then they may be more likely to be happy (hedonic). On the other hand, and as a security-related example, people might prefer to create short passwords or not waste time waiting for software updates to complete (preference), but the use of short passwords and declining software updates could make people's computer systems less secure (an objective list measure).

Rather than rely solely on a single definition of well-being and hence a single category of utilitarianism, those evaluating morality of actions may employ:

- **Pluralistic utilitarianism:** Pluralistic utilitarianism considers happiness, preference, objective lists, and other forms of benefits / harms in combination.

In moral considerations, a central focus is on the question, "what is the right decision to make?" However, the question "did we make the right decision?" is equally important, as it deals with questions of (retrospective) responsibility, redress, and retributive justice. When evaluating the moral quality of an action that has already happened, one view of consequentialism focuses on the actual outcomes regardless of what the likely outcomes were prior to the action. A probabilistic view of consequentialism asks whether the action was likely to have produced a net positive outcome regardless of whether it actually did so. Under the former view, an action that would likely have produced net negative results but that did not is still a right action; under the latter view, the action is not right.

Relatedly, when considering what decision to make, direct action utilitarianism focuses on the outcome of an action. Rule utilitarianism focuses on whether the decision follows rules designed to maximize positive net outcomes. Under rule utilitarianism, an action that causes net harm is still right if it follows rules that, across all scenarios, produce the greatest net positive results. We designed the scenarios in Section 3 to highlight key points of consideration about benefits and harms in individual situations and not as vehicles to discuss generalizable rules for the field. Hence, in Section 5, we adopt a direct action utilitarian perspective.

Deontological Ethics. Deontological ethics focuses on the moral duties of a given moral actor, such as an individual or an institution. These duties are often specified as direct duties (obligations) against others[7] — i.e., what does one person (morally) owe others [33]? These duties are often specified in terms of justice, either as negative duties (refrain from doing harm)[8] or as positive duties to certain claims that others have. In modern rights-based theories, these duties correspond to (moral) rights of the moral patient to whom the duty is owed. For example, if one person has a right to privacy, others owe this person (the moral patient) a certain behavior associated with that right.[9]

A defining result of the duty- / rights-based approach of deontological ethics is the focus on the right intention to act. While consequentialist ethics are mainly concerned with the outcomes, deontological ethics ask whether the action is undertaken out of a consideration for one's moral duty, or by some other thought process. Only an action that is performed with the intention to discharge a moral duty is considered moral.[10] Kant as one of the main protagonists of deontological ethics distinguishes between acting morally (i.e., out of consideration for one's moral duty) and legally (e.g., out of consideration for an actual legal framework or out of fear of sanctions). For example, completing a human subjects review process, such as an IRB within the U.S., solely because it is a university requirement is not a moral act.

Deontological ethics differ widely in their justification of the respective duties. One historical example is Divine Command and the duty to fulfill God's will. Most famously, Kantian ethics derives the moral duties from the faculty of reason that human beings have: because we can reason about what to do and thus control our desires, we have the obligation to do so in order to become autonomous (giving ourselves the moral law). And, since we are all potentially autonomous, we have a duty to treat all other human beings as such, i.e., as "ends and never purely as means" in Kant's words. A modern version of this Kantian thought is a specific take on contractualism [47], which posits that we should act in a way that cannot be reasonably rejected by anyone. Natural rights theories, on the other hand, take the idea that human beings as moral actors have certain faculties and justify natural (i.e., unalienable) rights (and corresponding duties) from those faculties for all persons (e.g., John Locke's "life, liberty, and estate" or the Virginia Bill of Rights) [37].

Given the influence of Kantian deontological ethics on

[7] This is one aspect that differentiates deontological reasoning from consequentialism, which at most posits a general duty to be moral (i.e., produce the best outcomes).

[8] While consequentialism is also concerned about (overall) harm, it does not hold that there are specific duties to the single individuals not to harm them. Rather, it aggregates harms and benefits, such that one given individual might suffer considerable harm if the net benefit for others is positive.

[9] These rights are often spelled out in Hohfeldian claim rights and corresponding obligations. As in our scenarios, the moral agents (researchers, program committee members, and so on) may incur direct moral obligations, Hohfeld's distinction between claim rights and liberty rights is not important here.

[10] This is not to say that deontological ethics entirely disregards the outcomes of an action. Neo-Kantian versions like John Rawls' "difference principle" (unequal distribution of certain goods is just, as long as it also benefits the least well-off), for example, often add a consequentialist aspect to the otherwise deontological reasoning. However, also in these cases the primary factor is individual rights and the intention to discharge duties associated with these rights.

the Belmont Report [54] and the Principles of Biomedical Ethics [7], and hence on the Menlo Report [56] and (sometimes implicit) arguments within the computer security research community, we take a Kantian approach to our ethics analyses. In order to make deontological ethics more tangible for computer security ethics and to contrast it more sharply with consequentialist ethics, in this paper we make a somewhat simplified assumption that deontological ethics conducts moral evaluation in the form of (individual or collective) duties and corresponding (individual) rights, that are spelled out in (absolute) terms of right or wrong. For example, if it is a duty not to harm someone, then killing one person to save five other lives — as presented in the classical trolley problem — directly interferes with this duty (and the person's right not to be killed) and is therefore wrong, no matter what.[11]

In contrast, consequentialist ethics, as exemplified by utilitarianism, conducts moral evaluation in the form of overall well-being (net utility), which allows comparative evaluations. The state in which only one person is dead will thus typically be a better state than the state in which five people are dead. While deontological ethics focuses on the intention (to discharge one's moral duties and to honor the rights of others to be treated in a certain way), consequentialist ethics focuses on the outcome of an action, policy, social practice, and more.

4.2 Other Ethical Frameworks

While there are other ethical frameworks, for the purpose of this paper we focus on consequentialist and deontological accounts (Section 4.1). These frameworks already have a strong presence within the security community, e.g., in the Menlo Report [56]. Indeed, elements of these frameworks are (at least implicitly) present whenever a researcher weighs benefits and harms or considers human rights. By construction (Section 3.1), these are also not only the easiest frameworks to apply in the scenarios we present, but also contrast each other (at least in the simplified scenarios that we have established). Nonetheless, we wanted to give a brief overview of some other ethical frameworks that seek to answer the question: What does it mean to act morally?

Virtue Ethics. Virtue ethics focuses more on how actors should *be*, i.e., a (morally) virtuous person, than on how they should *act*, since from being a virtuous person, (morally) virtuous actions will follow. In the Western world, the virtue ethical tradition can be traced back to Plato and Aristotle, who argued that people should cultivate and internalize virtuous moral character traits. Once internalized, a virtuous person would act virtuously — would choose the right action — by habit and by instinct. While "by habit and by instinct" at least for some authors necessitates immediacy, virtuous persons will still often use practical wisdom ("phronesis") in order to assess new situations and how they relate to internalized virtues. As this practical wisdom is itself a virtue, employing it in the right fashion and right instances also shows a virtuous person. While (in this Aristotelian sense) truly virtuous persons act on their internalized moral dispositions, they may also take time to thoughtfully consider the available actions, the expected outcomes, their responsibilities, and more, before deciding upon which action to take. Elements of virtue ethics manifests in the security community, for example, through the internalization of approaches to ethics after repeatedly considering ethics and security research over time, or whenever a program committee asks the question, "did the authors realize that they might have caused harm?"

While much of ethics / moral philosophy centers Western history, the internalization of virtuous character traits is present in numerous other traditions around the world, such as the *yamas* (external ethical practices) and *niyamas* (internal ethical practices) of the eight-fold Yogic traditions and the striving for *mushin* (an empty mind without motives or ego) in Japanese tradition. Likewise, and relatedly, Buddhist ethics and Confucian ethics also center virtues [44].

Discourse Ethics. Unlike consequentialist, deontological, and virtue ethics, which focus on the actions or actors, discourse ethics center a *process* for moral decision making, namely an *idealized moral discourse* that is egalitarian, inclusive, and principally interminable. Under discourse ethics, the belief is that whenever this idealized moral discourse nears a consensus, then this consensus is as close as we will ever get to moral truth. Discourse ethics provides a process for determining the morally right action even when, *a priori*, there is disagreement in what is morally right. In order to make this framework operationalizable, a key consideration has to be how to transform this justificatory claim about moral truth in a principally boundless moral community into the moral status of a real discourse within the computer security community. Another question is what constitutes "idealized moral discourse' within the realm of computer security. For example, could discussions within ethics review committees be regarded as an — albeit very specific and expert — idealized moral discourse?

Principlism Ethics. Principlism ethics believes that there are relatively uncontroversial principles upon which most moral theories converge, i.e., that there are a few principles central to the other frameworks combined. These principles should be used as a starting point to assess the morality of actions, institutions, policies, and more. One example is the four principles of biomedical ethics from Beauchamp and Childress [7]: respect for autonomy, beneficence, non-maleficence, and justice. Also published in the same era (1970s) are the principles in the Belmont Report [54], which focuses on the protection of human subjects in research. The Belmont principles are: respect for persons, beneficence, and justice.

[11] Of course, there are also pro-tanto-duties, which only hold as long as no more important or pressing duty surfaces. In order to contrast the two traditions more sharply, however, we ignore these in this paper.

The Belmont [54] and the Beauchamp and Childress [7] principles derive from both consequentialist and deontological ethics. Under principlism ethics, when faced with a moral decision, the actors should evaluate their courses of action with respect to these identified, universal principles. When conflicts arise between the application of different principles, e.g., between a consequentialist-derived principle and a deontological-derived principle, a process must be used to resolve those conflicts. The Menlo Report articulates the application of the Belmont Report's principles to the computing research field [56]. To make use of principlism for computer security scenarios, one must specify these principles at a concrete level and also formulate rules of priority in cases of value conflicts.

Emancipatory Ethics. Emancipatory ethics is the kind of ethics that Critical Theory would call its own, if Critical Theory entertained a dedicated ethics.[12] This kind of ethics is our umbrella term for more specific ethical enterprises, such as Ideology Critique, or certain versions of Care Ethics and Feminist Ethics. Common to all of these is the focus on the self-emancipation of individuals or groups that are in some way oppressed or marginalized. In order to achieve this outcome, emancipatory ethics centers the emancipated life — a life in which everyone can know and pursue their own "true interests" [23]. Emancipatory ethics therefore does not provide a framework to distinguish (moral) right from wrong, and not even an account of the conditions necessary such that those affected may be able to do so themselves. Rather, it criticizes current conditions as detrimental to a self-reflection on the affected's "true interest" and the individual and social transformations necessary to realize these interests [28]. More concretely, it analyzes social structures and relations in order to uncover and understand inequities, power structures, oppression, and discrimination, which will prevent those affected by these structures to "truly" self-reflect on their interests and associated moral obligations and to express them freely [30]. Under emancipatory ethics, there is no right moral action. Rather, it points to inequities, power structures, oppression, and more, not because of them being morally wrong, but because they inhibit the reflective moral process of those affected by those structures. Moreover, the belief is that those in power are often as tied by these structures as the oppressed, but with more to lose. Hence, under emancipatory ethics, it is essential to (1) highlight the results of the above-mentioned analyses and (2) thereby enable those subjected to these power asymmetries to contest them and demand change. An example of (2) would be including those with less power in the process of determining what is right, or inviting them to lead that process. As another example of the application of emancipatory ethics, if consequentialist- or deontological-like analyses are performed, then inequities, power structures, oppression, and discrimination must be considered centrally.

5 Analysis of Scenarios A, B, and C

We now turn to using consequentialist and deontological ethics (Section 4) to analyze the scenarios in Section 3. We encourage readers to review Scenarios A, B, and C first and consider what decisions they would make before reading our analyses. If readers wish, they may complete an online Google Form with their decisions (link available at https://securityethics.cs.washington.edu) and, upon doing so, see how others chose to respond.[13]

5.1 Analysis of Scenario A (Medical Device Vulnerability)

Here we consider the medical device vulnerability scenario from Section 3.2.

Consequentialist Ethics. Physical health in this scenario is an objective measure; if a patient chooses to remove a device or chooses not to obtain one because of a known vulnerability, then they would have a shorter life expectancy. Psychological health in this scenario is also an objective measure; if a patient knows about the vulnerability and still chooses to keep or get the implant, then they could live in fear of a security incident even though the likelihood of an incident is zero (by scenario construction).[14]

From a hedonic perspective, the knowledge that one has a shorter life expectancy (if they do not have the device) or the fear of a security incident (if they have the device) could lead to decreased happiness. In addition, the fact that removing or not opting for the device will result in ten years less of potential happiness may also significantly decrease overall happiness. From a preference utilitarian perspective, under the assumption that most patients would prefer not to learn about the vulnerability, then not disclosing the vulnerability would maximize the ability of people to live by their preference.

Hence, the morally correct decision is to *not disclose the vulnerability*.

Deontological Ethics. Under deontological ethics, the researchers have a duty to respect people's right to informed consent and the right to self-agency. In the medical context, this right to informed consent manifests (for example) as warnings in TV advertisements for medicines. These are fundamental human rights, and not disclosing the vulnerability

[12] The reluctance of many proponents of Critical Theory to talk about ethics or "an" ethics is often due to the perceived function of ethics to "tell others what is right or wrong". This goes directly against the emancipatory endeavor of Critical Theory to further "true" (ethical) self-reflection [29].

[13] The Google Form is anonymous — it requires Google authentication but does not reveal any identifiers to the authors of this paper. When interpreting the results of this form, we caution that no mechanisms, other than Google authentication, are used to protect against the use of different Google accounts to vote multiple times.

[14] In the real world, decision-makers must also consider family members, loved ones, and other stakeholders; we focus on patients for expositional simplicity.

would violate those rights. Hence, the morally correct decision is to *disclose the vulnerability*. This conclusion is correct even if most people would have preferred not to know about the vulnerability.

Informed by the Real World, Not Real. As discussed in Section 3, although real-world observations and experiences informed our scenario designs, there are gaps between our scenarios and what one might encounter in the real-world. Rather than choose from one of only two provided (binary) options, the researchers might, for example, choose to involve others in the decision-making process or cede the decision responsibility to another entity entirely. In the U.S., the FDA — not the researchers — could make or strongly contribute to the decision on whether to disclose the vulnerability to the public. Should they choose to disclose the vulnerability, they might work with healthcare providers to thoughtfully and conscientiously craft the message, thereby reducing patient alarm. Given the medical and security contexts, the decision-makers might leverage the Principles of Biomedical Ethics [7] and the Menlo Report [56]. Thus, even if the decision-makers do not solely rely on consequentialist or deontological analyses, and indeed consequentialist and deontological ethics both have limitations, consequentialist and deontological thinking may be part of the final decision-making process.

In Scenario A, we made the assumption that there is *zero* likelihood of the vulnerability being exploited regardless of whether or not the vulnerability is made public. We could have instead provided probability distributions for the likelihood of exploitation both if the vulnerability is made public and if it is not and then, for our consequentialist analysis, we could have calculated the likely overall benefits and harms for each decision. From a deontological perspective, if the public vulnerability disclosure would result in *more* people's devices being compromised than would be the case if there was not a public disclosure, then we would need to consider the negative impact of the public disclosure on those people's rights. By fixing the exploit probability at zero, our work is able to focus on comparing and contrasting the different ethical traditions rather than providing lengthy empirical analyses.

The above discussion points to another challenge with ethical decision making in the real world: uncertainty. In the real-world, a decision-maker might encounter questions that they cannot precisely answer, such as: Do all potential adversaries already know about the vulnerability? If not, then the public disclosure of the vulnerability might increase the exploit probability. On the other hand, if so, then the public disclosure of the vulnerability might not increase the exploit probability. That is unless, for example, the public disclosure of the vulnerability results in adversaries being more comfortable using their knowledge of the vulnerability. Would they be more comfortable? Or, supposing that adversaries do not already know about the vulnerability, what is the likelihood that they might discover the vulnerability themselves?

5.2 Analysis of Scenario B (Studying Immorally Obtained Data)

We now turn to analyzing the scenario in Section 3.3.

Consequentialist Ethics. Being able to find a job that one is qualified for is an objective measure of well-being in this scenario. The research has the potential to uncover biases or attack capabilities that can limit people's ability to find jobs. By proposing mechanisms to mitigate these biases or vulnerabilities, the research output can improve the ability of people to find such jobs.[15] Thus, from an objective list utilitarian perspective, the benefits of studying the data is high. Further, the data is already "public" and hence harm to job applicants has already happened. Further, the number of people harmed by the theft and release of the data is comparatively small compared to the one hundred-fold prediction of future use. Thus, from an objective list utilitarian perspective, the morally correct decision is *to study* the data.

A hedonic or preference utilitarianist would, respectively, observe that analyzing the data could degrade the happiness of the people whose data was stolen and would also prevent them to live by their preference, if they would prefer that the data not be studied. However, the number of people who would benefit (in both happiness and the ability to live by their own preferences) after the data is studied is far greater. Hence, even with the hedonic and preference utilitarianist frameworks, the morally correct decision is *to study* the data.

Deontological Ethics. Taking a Kantian deontological view, we observe that people have inalienable rights, including agency and privacy. Those rights extend to data intended to be private, whether it is private or not. Further, even if the right to privacy did not extend to adversarially-released data after it becomes public, in this scenario, the victims of the data breach have explicitly requested that their data be deleted everywhere. In order to do their research, the researchers would need to retain a copy of the data, thereby disrespecting the request to delete all data copies. They would also need to retain a copy *after* their research is complete in case their results are challenged, e.g., by Company B.

One might observe that future job applicants have a right to be treated fairly during the job application process, that the research results could result in a more fair AI system, and hence ultimately that the research would result in greater respect for the rights of future job applicants. However, under Kantian deontological ethics, individuals enjoy dignity. In Kant's own terms, this means that individuals may *not only* be treated as a means, but also always as *an end in itself*. Violating privacy rights in order to study the data (and prevent future harm) amounts to treating those whose data is studied solely as means to a different end, and is therefore wrong.

[15] Recall from Section 3.3 that addressing biases will improve the ability of some people to find a job (those impacted by biases) but, to simplify the analysis, will not negatively impact the ability of other people to find a job.

One might ask whether it would be appropriate to contact victims of the data breach and ask if their data can be retained and used for research — i.e., to obtain those victims' informed consent. This scenario does not present that option to the researchers. However, even if it did, under a deontological perspective, the act of asking a victim for informed consent in this scenario requires using the stolen data (to obtain victim identity or contact information); using the stolen data in this way is already a violation of privacy. Further, the act of contacting the victims could have unknown consequences.

Therefore, under Kantian deontological ethics, the stolen data should *not be studied*.

Informed by the Real World, Not Real. As with Scenario A, this scenario is informed by real-world experiences and observations but is not real. Researchers in the real world might have a mandatory first step prior to analyzing the data, e.g., if the researchers are in the U.S., they should work with their institution's IRB. The IRB would leverage the principles in the Belmont Report [54], which itself includes both consequentialist and deontological reasoning. The researchers may also seek input from others. For example, they may seek input from AI and security ethics experts, who might then also reference consequentialist, deontological, and other ethical frameworks. The researchers might also seek input from populations impacted by the study or non-study of the data, including representatives of people impacted by the data breach and representatives of people who could be harmed by the perpetuation of biases in Company B's AI system. Moreover, the researchers might offer these groups the option to lead the decision on whether to study the data.

5.3 Analysis of Scenario C (Inadvertent Data "Disclosure")

We now turn to the scenario in Section 3.4.

Consequentialist Ethics. The consequentialist must weight harms against benefits. There are harms to authors if the employee of Company C discloses the vulnerability to Company C — the authors will not be able to disclose at their preferred time (perference utilitarianism) and may be unhappy (hedonic utilitarianism) and may have their careers or other aspects of their lives negatively impacted if the early disclosure to Company C limits their impact or ability to publish (career advancement could be a measure of well-being per objective list utilitarianism).

However, the harms to Company C's users if the employee does not disclose the vulnerability to Company C is much greater — without early disclosure, Company C will not be able to protect their users and, as a result, millions of people around the world could be significantly harmed.

Hence, the morally correct action is for the employee *to disclose* the vulnerability internally to Company C.

Deontological Ethics. The employee of Company C may feel a sense of duty to their company and to their company's users. However, program committee members also have a duty to respect the autonomy of authors and a duty to respect the confidentiality of the peer review process. Moreover, the employee of Company C agreed to respect this duty when they joined the program committee, as did Company C's leadership team when they granted the employee permission to join the program committee. Moreover, from a Kantian reasoning, the employee could not form a maxim that allowed breaking the confidentiality promise, as otherwise the peer review process and the institution of program committees would not be possible.

Thus, the morally correct thing for the Company C employee to do is respect the rights of the authors and the confidenitality of the review process and *not disclose the vulnerability* to Company C.

Informed by the Real World, Not Real. In a real-world scenario, the employee of Company C might not make the decision on their own. For example, rather than decide between the two options we presented, they might first reach out to the program chairs and ask them to give advice or render a decision. The program chairs might then explore questions such as: should they reach out to the paper's authors, asking for more information about the disclosure timeline? If Company C's employee assesses the harms of the vulnerability as significant, should the program chairs ask the authors to disclose to Company C right away? What are the impacts on the scientific peer review process if the program chairs ask the authors to disclose to Company C? Would the authors feel compelled to grant permission because they want their paper to be accepted even if granting permission is not in their best interest? Since not all companies have members on the program committee, is it morally right to give this company (and their users) advance notice of a vulnerability (even with author permission) solely because an employee of their company is on the program committee?

6 Discussion

6.1 Reflection on Analyses

We begin by reflecting upon our analyses and summarizing key points and observations. We are not claiming that all these reflections and observations are novel and, indeed, many ideas herein are likely familiar to those with expertise in ethics; see Section 4 for a survey of resources on ethics / moral philosophy. We offer these reflections and observations because we hope that they can serve to further our community's collective thoughts and perspectives on ethics and computer security.

Different Frameworks Can Lead to Different Conclusions. For some moral questions, different ethical frameworks lead to *different* conclusions regarding what is right and wrong.

Different Frameworks Can Lead to the Same Conclusion. For other moral questions, different ethical frameworks lead to the *same* conclusion regarding what is right and wrong.

A Framework Can Fail to Reach a Conclusion. We intentionally designed our scenarios to be "decisive", per the goals in Section 3.1; real-world scenarios may *not* be decisive and may *not* lead to conclusive decisions under either the consequentialist or deontological frameworks. Also, it could be the case that under a framework a certain action is morally permitted, i.e., not necessarily required but also not forbidden.

Ethical Frameworks Can Provide Tools for Discussion. What should one do when there are differences of opinion or lack of clarity into what constitutes the right decision? Here is where the tools — the frameworks — from ethics / moral philosophy can help. In short, they can help decision-makers thoughtfully, methodically, and articulately analyze moral questions.

In discussions of right or wrong, when there is disagreement, we suggest first surfacing the communicants' underlying values and their frameworks of consideration. Simply knowing that another communicant is centering different values and a different framework may help further a collaborative discussion.

Ethical Frameworks Can Provide Tools for Thought. In this work, we primarily consider consequentialist and deontological ethics. Both of these frameworks have limitations, and we are *not* advocating for strict adherence to either of them. In fact, it is not uncommon for people — including modern ethicists — to include elements of multiple frameworks (consequentialist, deontological, and other) as they reason through decisions. Within the security research community, the Menlo Report [56] includes both consequentialist and deontological elements, for example.

On the one hand, the observation above might call into question the value of articulating ethical frameworks in the first place: if people are not strictly consequentialist or deontological, what value is there in exploring scenarios from strict consequentialist or deontological perspectives? We argue that precise analyses of scenarios under different perspectives can help the decision-maker in multiple ways. At a minimum, precise thinking via the ethical frameworks can help slow the decision-making process and encourage thoughtful reflection and contemplation. Additionally, the frameworks can help decision-makers identify which parts of arguments they agree with and which parts they do not and, by doing so, help the decision-maker better articulate their own arguments, even if their arguments are neither consequentialist nor deontological.

Sometimes the Morally Correct Action is Not in the Best Interest of the Decision-Maker. In Scenarios A, B, and C, we tried to minimize the impact of either decision on the decision-makers themselves. Thus, the decision-makers could focus on the impacts on and rights of others. In the real world, a decision-maker's decision might also impact themselves (e.g., a researcher might desire a publication, and the decision on how to proceed might impact their ability to publish). We explore such situations in Scenarios D* and F in the full version of this paper [34]. In short, sometimes the morally right decision might *not* be the decision that seems to be in the best interest of the decision-maker.

Shifting Morality Earlier. Our scenarios all feature moral questions for decision-makers. However, one might ask (not just for our scenarios, but for the field) how to shift questions of morality earlier, such that the scenarios we consider (or the real world encounters) do not come up. In a trolley problem, an example of "shifting earlier" might be to ensure that all trolleys have better, more resilient brakes. For Scenario A, code escrow might enable the patching of devices even after the manufacturer ceases operation. For Scenario B, the researchers would not have needed to study the data if the underlying AI algorithms were already unbiased and secure. For Scenario C, the situation could be mitigated if the conference had pre-specified rules for such situations or if the conference required disclosure before submission; of course, whether such a rule should be in place raises its own ethical questions, as exhibited (for example) by the role of legal threats in Scenarios D* in the full version of this paper [34].

On Uncertainty. Within the computer security field, there are significant elements of uncertainty. One challenging element of uncertainty surrounds that of the adversary. It is often not known to decision makers when or if adversaries might manifest. Further, precise adversarial capabilities are seldom known to decision makers in advance of their manifestation. Additionally, there is generally uncertainty regarding what unknown vulnerabilities a system might have. For our core scenarios, we aimed to reduce uncertainty through the concrete, precise description of outcomes for each decision option. A challenge for the computer security research community, and an opportunity for ethics and moral philosophy researchers, is to formulate frameworks and candidate approaches for ethical decision making in the presence of uncertainty, including uncertainty about adversaries, adversarial actions, and unknown vulnerabilities. (It is because of the challenges imposed by uncertainties that we explicitly incorporate uncertainty into Scenarios D* in the full version of this paper [34].)

On the Role of Details and Time. Ethical assessments not only depend on moral arguments (e.g., whether we should maximize utility understood as X, whether one has a right to Y or a duty towards person Z), but also on non-moral considerations about the specifics of a given context. Therefore, it is difficult to provide ethically sound judgments that persist over time (as the field of computer security evolves and learns more), as well as across the nuances of different scenarios. On the latter, we refer to the D* and E* scenario sequences in the full version of this paper [34] and the impacts of scenario changes on what one might perceive as morally right or wrong. Likewise, one might consider the impact of changes to the specifics of Scenarios A, B, and C, e.g., if impacted job applicants did not explicitly request the deletion of stolen data in Scenario B or if the program chairs did not explicitly mandate confidentiality in Scenario C. Further, our field's understanding of harms and adverse impacts evolve over time, e.g., as the field's knowledge of adversarial capabilities or the (possibly) harmful consequences of a research method matures. Hence, what might be seen as morally right or permissible at one point in time may, at a later date, be seen as neither morally right nor permissible, and vice versa. Collectively, these are *not* reasons not to strive for detailed, specific rules (or at least guidelines) for ethically sound behavior. Rather, we suggest that the specific contexts are among the factors that the security community must consider if it were to do so. We argue that the security community should therefore formulate the rules as specific as possible *and* as general as necessary in order to abstract from the details of a given context. A first consideration here could be that (in much the same way as in legal texts) the impartial nature of morality requires the treatment of like cases alike. Therefore, guidelines should address types of cases, rather than single out individual cases, but with consideration both for when and where generalizability applies and when there may be challenges to generalization.

6.2 For Consideration

With the background of our results and our reflections, we now present a collection of considerations for members of the security research community.

For Decision-Makers. Decision-makers (researchers, program committees, others) should consider ethics *before* making decisions, rather than after. For certain moral dilemmas (e.g., Scenarios A, B, and C), it is possible to pick an outcome and then find the ethical framework that justifies that outcome. We do *not* argue for this practice. Instead, decision-makers should let the decision follow from a disinterested ethical analysis. Toward facilitating disinterested analyses, we encourage decision-makers to explicitly enumerate and articulate any interests that they might have in the results of the decision; such an articulation could be included as part of a positionality statement in a paper.

For Researchers Writing Papers. For researchers new to ethics, the Menlo Report [56] provides concrete guidance. The Menlo Report and other ethical frameworks can help researchers reach a conclusion about what is morally right.

This paper, we hope, can help researchers consider and discuss morality when there are differences of opinion or uncertainty regarding what to do. Because (1) we believe that the field can grow through the explicit articulation of ethical thought and (2) there can be differences in ethical perspectives and thought (as Section 5 shows), we encourage researchers to do more than just apply a single approach (consequentialist, deontological, the principles in the Menlo Report, or otherwise) and then act accordingly. Rather, we encourage researchers to conduct analyses under multiple ethical frameworks *and* include the reasoning for their decisions under the multiple frameworks in their paper submissions and publications. If the frameworks lead to the *same* conclusion, the inclusion of multiple arguments can strengthen the paper's ethics section and can serve as part of the growing foundation for ethical thought in the field. If different frameworks lead to *different* conclusions, and the authors proceed with what is considered morally right under one framework but morally wrong under another, then surfacing those different considerations and the final thought processes can be particularly valuable. For example, if papers start including analyses under multiple frameworks, then such analyses could become the norm and published analyses could become additional guides for future researchers.

To aid in the above, we propose a process that we call *ethics modeling*. This process builds on the Menlo Report [56] and other approaches for evaluating ethics [38] as well as on some approaches to threat modeling. Namely, we suggest that researchers first do a stakeholder analysis to identify all stakeholders potentially impacted by the decision, e.g., using methods from value sensitive design [21]. Then, for each stakeholder, we suggest explicitly identifying the assets that might be impacted by the possible decisions. Then, for each possible decision, for each stakeholder, and for each asset, enumerate the benefits / harms (consequentialist ethics) and the rights supported / violated (deontological ethics). The benefits / harms and rights analyses should consider situations in which no adversaries manifest and situations in which adversaries manifest. We call this process as *ethics modeling* because it combines elements of both ethical analyses and threat modeling.

We further encourage researchers to become familiar with ethical frameworks not deeply considered in this work. An example in the context of computer security and victims of intimate partner violence is care ethics, as considered in Section 6.2 of Tseng et al. [53].

For Program Committees Discussing Submissions. We encourage program committees and paper reviewers to become familiar with the different ethical frameworks. When questions of ethics arise in the review process, we encourage program committee discussions to explicitly reference not just *what* the discussants believe is morally right and wrong, but *why* they believe that. The latter — the why — can explicitly refer to analyses under one or more ethical frameworks. Further, we encourage reviewers to strive to infer what ethical framework or approaches the authors took, if any and if not explicitly articulated, and to consider that the authors may have centered a framework or approach that differs from that of (at least some of) the reviewers. The authors' approach might have been shaped by an academic environment or culture different from the reviewers' own, for example.

For the Community. We encourage the community at large to familiarize themselves with different ethical frameworks. Those community discussions could leverage the scenarios that we developed over the course of this research. For example, prior to reviewing papers, a program committee could, together, discuss the committee's perspective on the right decisions for the scenarios that we present. From preliminary conversations with members of our community, we believe that such discussions will not lead to a universal consensus. But we believe that the resulting conversations, and the points raised, would be helpful for those community members as they, for example, embark on reviewing papers with possible ethical concerns.

Additionally, we encourage continued community-wide conversations around infrastructure support for proactive, pre-reseach considerations of ethics and morality beyond what is traditionally covered by IRB. For example, the security community might draw inspiration from the Ethics and Society Review Board as implemented by Stanford HAI [8] as well as existing approaches for peer-review prior to the implementation of a research method, e.g., [11]. As with research efforts and program committee reviews, we believe that such evaluations would benefit from considerations under, or at least awareness of, multiple ethical frameworks.

For Educators. We encourage educators to include explicit discussions of ethics and ethical frameworks in their courses if they are not already doing so. Our Scenarios A, B, and C, by design, do not lead to obvious right and wrong answers. As a result, we have found that our scenarios are particularly conducive to conversations in classes. Educators are welcome to use our scenarios in their classes as well. A companion slide deck is available at https://securityethics.cs.washington.edu.

For Industry and Government. Although we have scoped our work to focus on computer security research, we believe that this work may also be of interest to those in industry, government, and other sectors. One concrete suggestion, as articulated by an anonymous reviewer and included with permission, is for the Internet Engineering Task Force (IETF) to consider requiring an "Ethical Considerations under Multiple Frameworks" section in each Internet-Draft, much like the present requirement of a "Security Considerations" section.

For Everyone. Creating ethical norms for computer security research is fundamentally challenging because different ethical frameworks can lead to different conclusions about right and wrong. We believe that a more achievable near-term goal is the creation of extensive sets of case studies (like our scenarios) that community members can discuss and learn from.

For Us. Although we are confident that our dilemmas in Scenarios A, B, and C are true dilemmas (per our criteria in Section 3.1, our validation methodology, and extensive iteration and discussion), this version of our paper does not report concrete data (as doing so is not the goal of this paper). Our ongoing work seeks to provide such concrete data across cultures and communities. We are additionally preparing other computer scenario descriptions, in the format of the scenarios in this paper, for community consideration. As we create these scenarios, we will add them to https://securityethics.cs.washington.edu/.

7 Conclusions

In this paper, we embark on a research collaboration spanning (1) ethics / moral philosophy and (2) computer security research. We develop criteria for computer security-themed trolley problems. We present three such trolley problems (Scenarios A, B, and C) and then evaluate those trolley problems under today's main ethical frameworks. Given the findings of our research, we reflect and offer considerations for the computer security research community.

Acknowledgements

This work was supported in part by the U.S. National Science Foundation under awards CNS-2205171 and CNS-2206865, the University of Washington Tech Policy Lab (which receives support from the William and Flora Hewlett Foundation, the John D. and Catherine T. MacArthur Foundation, Microsoft, and the Pierre and Pamela Omidyar Fund at the Silicon Valley Community Foundation), and gifts from Google, Meta, Qualcomm, and Woven Planet. We are grateful to everyone who contributed to this project. Thank you to all who offered comments, questions, insights, and conversations, hosted talks, and reviewed preliminary drafts, including Lujo Bauer, Hauke Behrendt, Dan Boneh, Kevin Butler, Aylin Caliskan, Inyoung Cheong, Lorrie Faith Cranor, Sauvik Das, Zakir Durumeric, Kevin Fu, Alex Gantman, Gennie Gebhart, Kurt Hugenberg, Umar Iqbal, Apu Kapadia, Erin Kenneally, David Kohlbrenner, Seth Kohno, Phil Levis, Rachel McAmis, Alexandra Michael, Bryan Parno, Elissa Redmiles, Katharina Reinecke, Franziska Roesner, Stefan Savage, Stuart Schechter, Sudheesh Singanamalla, Patrick Traynor, Emily Tseng, and Miranda Wei. We thank the anonymous USENIX Security 2023 reviewers for their insightful feedback, comments, and suggestions. We also sincerely thank all attendees of past presentations about this work.

References

[1] ACM. ACM code of ethics and professional conduct, 2018. https://www.acm.org/code-of-ethics.

[2] Eytan Adar. User 4xxxxx9: Anonymizing query logs. In *Proc of Query Log Analysis Workshop, International Conference on World Wide Web*, 2007.

[3] Henrik Ahlenius and Torbjörn Tännsjö. Chinese and Westerners respond differently to the trolley dilemmas. *Journal of Cognition and Culture*, 12(3-4), 2012.

[4] G.E.M. Anscombe. Modern moral philosophy. *Philosophy*, 33(124), 1958.

[5] John W. Ayers, Theodore L. Caputi, Camille Nebeker, and Mark Dredze. Don't quote me: Reverse identification of research participants in social media studies. *npj Digital Medicine*, 2018.

[6] Julian Baggini and Peter S. Fosl. *The Ethics Toolkit: A Compendium of Ethical Concepts and Methods*. Blackwell Publishing, 2007.

[7] Tom L. Beauchamp and James F. Childress. *Principles of Biomedical Ethics*. Oxford University Press, 8 edition, 2019. First edition published in 1979.

[8] Michael S. Bernstein, Margaret Levi, David Magnus, Betsy A. Rajala, Debra Satz, and Quinn Waeiss. Ethics and society review: Ethics reflection as a precondition to research funding. *PNAS*, 118(52), 2021.

[9] Amy Bruckman. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology*, 4(3), 2002.

[10] Ben Burgess, Avi Ginsberg, Edward W Felten, and Shaanan Cohney. Watching the watchers: Bias and vulnerability in remote proctoring software. In *USENIX Security*, 2022.

[11] Center for Open Science, 2023. https://www.cos.io/initiatives/registered-reports.

[12] Paul Conway and Bertram Gawronski. Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104(2), 2013.

[13] Paul Conway, Jacob Goldstein-Greenwood, David Polacek, and Joshua D Greene. Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition*, 179, 2018.

[14] John Deigh. *An Introduction to Ethics*. Cambridge University Press, 2010.

[15] Fred Donovan. FDA unveils MITRE's medical device security playbook, 2018. https://healthitsecurity.com/news/fda-unveils-mitres-medical-device-security-playbook.

[16] Julia Driver. *Ethics: The Fundamentals*. Blackwell Publishing, 2006.

[17] Brianna Dym and Casey Fiesler. Ethical and privacy considerations for research using online fandom data. *Fan Studies Methodologies*, 33, 2020.

[18] Casey Fiesler and Nicholas Proferes. "Participant" perceptions of Twitter research ethics. *Social Media + Society*, 4(1), 2018.

[19] Luciano Floridi. *The Cambridge Handbook of Information and Computer Ethics*. Cambridge University Press, 2010.

[20] Philippa Foot. *Moral Dilemmas: And Other Topics in Moral Philosophy*. Oxford University Press UK, 2002.

[21] Batya Friedman and David G Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, 2019.

[22] Bernard Gert and Joshua Gert. The definition of morality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition, 2020.

[23] Raymond Geuss et al. *The Idea of a Critical Theory: Habermas and the Frankfurt School*. Cambridge University Press, 1981.

[24] Natalie Gold, Andrew M. Colman, and Briony D. Pulford. Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making*, 9(1), 2014.

[25] J. Habermas and C. Cronin. *Justification and Application: Remarks on Discourse Ethics*. Studies in Contemporary German. MIT Press, 1994.

[26] Daniel Halperin, Thomas S. Heydt-Benjamin, Benjamin Ransford, Shane S. Clark, Benessa Defend, Will Morgan, Kevin Fu, Tadayoshi Kohno, and William H. Maisel. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *IEEE Symposium on Security and Privacy*, May 2008.

[27] Richard Mervyn Hare. *Moral thinking: Its levels, Method, and Point*. Oxford: Clarendon Press; New York: Oxford University Press, 1981.

[28] Axel Honneth. *Freedom's Right: The Social Foundations of Democratic Life*. Columbia University Press, 2014.

[29] Max Horkheimer. *Critical Theory: Selected Essays*. Continuum Publishing Corporation, 1975. Originally published in 1937.

[30] Max Horkheimer and Theodor W. Adorno. Dialektik der aufklärung. *Philosophische Fragmente*, 14, 1947.

[31] IEEE. IEEE code of ethics, 2020. https://www.ieee.org/about/corporate/governance/p7-8.html.

[32] Ron Iphofen. *Handbook of Research Ethics and Scientific Integrity*. Springer, 2020.

[33] Immanuel Kant. Grundlegung zur metaphysik der sitten [groundwork of the metaphysics of morals]. *Riga, Latvia: JF Hartknoch*, 1785.

[34] Tadayoshi Kohno, Yasemin Acar, and Wulf Loh. Ethical frameworks and computer security trolley problems: Foundations for conversations, 2023. Available online at https://securityethics.cs.washington.edu and at https://arxiv.org/abs/2302.14326.

[35] Daniel B. Kramer and Kevin Fu. Cybersecurity concerns and medical devices: Lessons from a pacemaker advisory. *JAMA*, 318(21), 2017.

[36] Michèle Lamont, Laura Adler, Bo Yun Park, and Xin Xiang. Bridging cultural sociology and cognitive psychology in three contemporary research programmes. *Nature Human Behaviour*, 1(12), 2017.

[37] John Locke. *Locke: Two Treatises of Government*. Cambridge University Press, 1988. Originally published in 1689.

[38] Arne Manzeschke, Karsten Weber, Elisabeth Rother, and Heiner Fangerau. *Ethical Questions in the Area of Age Appropriate Assisting Systems*. 2015.

[39] Annette Markham. Fabrication as ethical practice. *Information, Communication & Society*, 15(3), 2012.

[40] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy*, 2008.

[41] NSPE. NSPE code of ethics for engineers, 2019. https://www.nspe.org/resources/ethics/code-ethics/.

[42] Derek Parfit. *Reasons and Persons*. OUP Oxford, 1984.

[43] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2), 2021.

[44] Michael Puett and Christine Gross-Loh. *The Path: What Chinese Philosophers Can Teach Us About the Good Life*. Simon and Schuster, 2016.

[45] Michael J. Quinn. *Ethics for the Information Age*. Pearson, 7 edition, 2017.

[46] Benjamin Ransford, Daniel B. Kramer, Denis Foo Kune, Julio Auto de Medeiros, Chen Yan, Wenyuan Xu, Thomas Crawford, and Kevin Fu. Cybersecurity and medical devices: A practical guide for cardiac electrophysiologists. *Pacing and Clinical Electrophysiology*, 40(8), 2017.

[47] Thomas M Scanlon. *What We Owe to Each Other*. Harvard University Press, 2000.

[48] Stanford University, Philosophy Department, Metaphysics Research Lab. Stanford encyclopedia of philosophy, 2022. https://plato.stanford.edu/index.html. Principal editors: Edward N. Zalta, Uri Nodelman; associate editors: Colin Allen, Hannah Kim, Paul Oppenheimer; assistant editors: Emma Pease, Lauren Thomas, Jesse Alama.

[49] Eliza Strickland and Mark Harris. Their bionic eyes are now obsolete and unsupported. *IEEE Spectrum*, February 2022.

[50] John Stuart Mill. Utilitarianism. *Parker, Son, and Bourn, London*, 1863.

[51] Latanya Sweeney. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics*, 25(2–3), 1977.

[52] Daniel R Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R Beresford. Ethical issues in research using datasets of illicit origin. In *IMC*, 2017.

[53] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. Care infrastructures for digital security in intimate partner violence. In *ACM CHI*, 2022.

[54] U.S. Department of Health, Education, and Welfare. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research, April 1979. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

[55] U.S. Food and Drug Administration. Cybersecurity, 2022. https://www.fda.gov/medical-devices/digital-health-center-excellence/cybersecurity.

[56] U.S. Homeland Security. The Menlo report: Ethical principles guiding information and communication technology research, August 2012. https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.

[57] Shannon Vallor, Irina Raicu, and Brian Green. Technology and engineering practice: Ethical lenses to look through, 2020. https://www.scu.edu/ethics-in-technology-practice/ethical-lenses/; from Ethics in Technology Practice, the Markkula Center for Applied Ethics at Santa Clara University, https://www.scu.edu/ethics/.

[58] Sheridan Wall and Hilke Schellmann. LinkedIn's job-matching AI was biased. The company's solution? More AI, June 2021. https://www.technologyreview.com/2021/06/23/1026825/linkedin-ai-bias-ziprecruiter-monster-artificial-intelligence/.

[59] Shoko Yamamoto and Masaki Yuki. What causes cross-cultural differences in reactions to the trolley problem? a cross-cultural study on the roles of relational mobility and reputation expectation. *The Japanese Journal of Social Psychology*, 2020.

[60] Michael Zimmer. Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media + Society*, 4(2), 2018.