Author accepted manuscript (AAM) of Lechterman, Theodore, Bradley Strawser, and David Whetham. "Ethical Safeguards for Sales of Weaponizable Technology in Advance: A Case Study." *Business and Professional Ethics Journal*, 2025. https://doi.org/10.5840/bpej2025324164.

Ethical Safeguards for Sales of Weaponizable Technology: A Case Study

Theodore M. Lechterman
IE University
Madrid, Spain
tlechterman@faculty.ie.edu

Bradley J. Strawser
Naval Post Graduate School
Monterey, CA, USA
bjstraws@nps.edu

David Whetham
King's College London
London, UK
david.whetham@kcl.ac.uk

Abstract

This article presents a case study in how sellers of weaponizable technology can develop safeguards to mitigate risks of misuse by end users. In 2020, the authors were approached by a defense technology start-up whose core product offering was weaponizable drones. The start-up sought guidance in designing terms of sale and service that would ensure responsible usage of this technology. Combining elements from just war theory, international humanitarian law, and the theory of responsibility, we developed a novel, systematic framework for reducing risks of misuse on the basis of precise principles and objective metrics. Although designed for a specific use case, we believe aspects of this framework are portable to a wide range of scenarios. We share it to demonstrate proof of concept and stimulate further work on integrating ethical considerations into the business of weapons and dual-use technology.

Acknowledgments

The authors gratefully acknowledge Teresa Allen for research assistance and manuscript preparation, Linda Eggert, Ryan Jenkins, and Michael Skerker for comments on earlier drafts, and Lisa Strutz for data visualization and project management.

Funding Sources

Research support was provided by the Mercatus Center at George Mason University, Award AIDG983718408.

Competing Interests Statement

As discussed directly in the manuscript, this work is a case study of an organization for whom the authors served as advisors. The subject organization dissolved in 2021, releasing the authors from any confidentiality requirements. Prior to its dissolution, the organization expressly granted the authors unrestricted permission to publish these results without input or review. The authors serve as independent contractors for Compass Ethics, an ethics consultancy.

Keywords: Ethics by design, weapons technology, dual-use technology, just war theory, ethics of sales, due diligence

Introduction

Powerful technology has the potential to save and improve lives. In the wrong hands, however, it can cause great harm. Numerous business ethicists propose that corporations have duties of due diligence to mitigate the human rights impacts of their operations and may be complicit in human rights failures that are causally connected to their work (e.g., Wettstein 2010; Fasterling and Demuijnck 2013). The inherent hazards of weapons or dual-use technology may make these duties especially stringent (Skerker, Lechterman, and Strawser n.d.). While one might assume that it is the primary responsibility of governments to regulate the terms of sale and usage of these products, regulatory voids are common with emerging technologies whose capabilities may not be fully known or knowable (Short 2013, 26-28). Moreover, companies may be given significant discretion over how to interpret broad regulatory principles (Coglianese and Mendelson 2010). In the absence of comprehensive regulatory oversight, technology companies may lack clear guidance on duties to prevent the misuse or abuse of their products.

However, attempts by companies to ensure responsible use run into difficult challenges. People disagree about the nature of misuse, how to assess it, and how to guard against it. Without a coherent framework for guiding decisions, judgments may be ad hoc and subjective. Even when these judgments are sound, the lack of a clear and settled process for reaching and implementing these judgments can cost an organization precious time and resources. A lack of a standardized process can also provoke backlash when stakeholders disagree with executives' decisions.

As companies have started to explore ways to systematize decisions about responsible use, many have been unable or unwilling to integrate ethics holistically.[1] Many, for example, have sought to place restrictions on usage or sales without first considering the purposes or capabilities of the product itself or the business model in which it is developed.[2] Not only have these approaches failed to achieve their intended goals, but they also result in incoherence across workstreams and conflict among stakeholders.

The proposal that follows was designed to overcome the foregoing challenges by offering a framework for systematizing and justifying decisions about product sales and use. It does so by offering a comprehensive approach, in which ethical guidelines support and unify each stage of the product life cycle. The guidelines that follow seek to connect (1) decisions about customer acquisitions, to (2) decisions about acceptable usage policies, to (3) decisions about monitoring compliance with those policies, to (4) decisions about preventing and responding to misuse, which also feed back into decisions about product design. These separate elements work together to deliver a powerful suite of tools for bringing greater ethical integrity to the business of weapons and dual-use technology.

This framework was originally designed by the authors at the request of a drone manufacturing start-up whose products could be used for reconnaissance and delivery of munitions in urban environments. The client believed that deploying such a product could serve morally important ends, by providing technical advantages to peacekeeping forces and dramatically reducing unintended casualties in combat. However, the client also recognized the significant possibilities of abuse and misuse of a product of this kind and sought our guidance on forecasting and mitigating

[1] See Winner (2014) for a discussion of "technological somnambulism"—i.e., our general tendency to welcome new technologies without giving much thought to the ethical risks involved.

[2] Consider Facebook's Oversight Board, which launched in 2020. The Board is designed to enhance public accountability and address threats such as hate speech and the spread of disinformation. However, as Ghosh (2019) argues, the approach fails to acknowledge ethical risks built into the platform's business model, which relies on selling personal data and curating content in the name of maximizing profits. Ghosh suggests a more earnest approach to addressing online threats would involve increased efforts at protecting consumer privacy, increased oversight of the platform's algorithms, and more.

these ethical risks. The company ultimately dissolved before going to market, freeing us from confidentiality requirements and providing a rare opportunity to publicize advanced and unfiltered policy work in this sensitive sector.[3] We share the framework here in the interest of demonstrating proof of concept and stimulating further work on integrating ethical considerations into business decisions regarding weapons and dual-use technology.

1. Establishing the Foundation of the Framework

A necessary first step toward any responsible use framework is to establish the baseline assumptions from which all decisions will stem. Suppose we start from the position that one should not sell arms to people, states, or agents. A consequence of this view is: if one is to sell arms, one is therefore seeking to make an exception to that normal prohibition.[4] Just War Theory (JWT)—also known as the Just War Tradition to acknowledge the way that its principles have been interpreted and applied to reflect the changing character of conflict over the past two millennia—provides us with a central starting point in evaluating the conduct of the exception (Aquinas, 2002; Augustine, 2003; Walzer, 1977; Whetham, 2010; McMahan, 2011; Strawser, 2023).

1.1. Just War Framework

We have drawn from JWT to frame much of our thinking in this area, as its ubiquitous nature provides "a common language for discussing and debating the rights and wrongs of conflict" (Whetham, 65). One core principle of JWT is the principle of discrimination: in war, if and only if the determination has been made that the act of war is indeed justified, one must nevertheless limit harms against those who have not made themselves liable to harm (i.e., non-combatants, civilians, etc.). Another is the principle of proportionality: if innocents would be unintentionally harmed in any way, that harm must be limited and proportionate to the legitimate goal being pursued. JWT also strictly prohibits "evils in themselves"—actions or tactics that are simply considered wrong, no matter the context.

As with selling arms, JWT starts from the assumption that war is *prima facie* wrong and, therefore, can only be justified as an exception to that normal position. Borrowing the foundational assumptions from JWT, we proposed the following set of principles—the somewhat canonical set referred to as the *jus ad bellum* criteria—as a baseline for determining when an exception to selling arms may be permitted (Strawser 2023; Cook 2004; McMahan 2005):

- **Just Cause:** the state in question must have a genuine justified cause as the purpose for which it intends to use weapons of war.
- **Just Intention:** the state must intend to use the weapons for the purpose of pursuing the just cause with an ultimate aim of restoring peace.
- **Last Resort:** the state must exhaust other non-violent means short of war to rectify or prosecute the just cause before resorting to war.

---

[3] The company faced COVID-19 pandemic-related disruptions and conflicts over leadership and strategy. For more context, see Kempf and Shapiro (2024).

[4] This is a familiar starting point in the literature on dual-use technology. It is grounded in what Miller (2018) refers to as the "No Means to Harm Principle," according to which it is morally impermissible to "avoidably and foreseeably (whether intentionally or unintentionally) provide others (directly or indirectly) with the means to intentionally (or negligently) do harm" (13). As Miller points out, though, exceptions to this principle can be made. E.g., although experts and the public may disagree about when the use of lethal force is permissible, it is generally agreed upon that military forces and police units may use lethal weapons to protect the public and themselves.

- **Reasonable Hope of Success:** the state must be able to plausibly succeed in its efforts to achieve the aims of the just cause, rather than simply wasting lives in a lost cause.
- **Legitimate Authority:** the state must be acting in accordance with the accepted political mechanisms of the state, representing, in some sense, the will of the people going to war.
- **Global Proportionality:** the war efforts must be worth the predicted costs, all things considered, when compared to the just cause for which the state is fighting.

1.2 Customary International Humanitarian Law

In establishing the framework, we also leaned heavily on elements of JWT principles that have been codified into international law. While some could argue that the JWT principles alone should structure the framework, relying entirely on JWT principles was impractical for the case at hand.[5] The framework was designed to aid a particular business in its customer selection and product usage policies. Concerns for legitimacy and feasibility tell in favor of linking the framework as much as possible to accepted and codified standards. International humanitarian law (IHL) provides this. IHL is a collection of rules and agreements designed to reign in and limit the harms of warfare, for humanitarian reasons, to as great an extent as possible, while acknowledging the reality that war will still sometimes be fought. It is the legal framework applicable to all cases of armed conflict, occupation, civil war, and any large-scale use of force or violence on behalf of, in concert with, or against a state or state-like-entity. The central set of agreements are known as the Geneva Conventions, which were first established in 1949, and have expanded several times through key Additional Protocols I, II, and III.

Several important international bodies and organizations safeguard these various legal instruments and interpretations that make up IHL and are considered authoritative upon them. One such organization is the International Committee of the Red Cross (ICRC). The ICRC maintains a database of the key provisions, rules, and expectations of practice that derive from across the breadth of Customary IHL.[6] The ICRC provides a condensed synthesis of these rules in a helpful summary "which sets out, as simply and briefly as possible, the fundamental rules which are the basis of these treaties and the law of armed conflicts as a whole" (de Preux 2023, 6). Those seven basic rules of IHL are as follows:

1. Persons *hors de combat* (outside of combat) and those who do not take a direct part in hostilities are entitled to respect for their lives and their moral and physical integrity. They shall in all circumstances be protected and treated humanely without any adverse distinction.
2. It is forbidden to kill or injure an enemy who surrenders or who is *hors de combat*.
3. The wounded and sick shall be collected and cared for by the party to the conflict which has them in its power.[7]
4. Captured combatants and civilians under the authority of an adverse party are entitled to respect for their lives, dignity, personal rights, and convictions. They shall be protected against all acts of violence and reprisals.
5. Everyone shall be entitled to benefit from fundamental judicial guarantees. No one shall be held responsible for an act he has not committed. No one shall be subjected to physical or

[5] We thank an anonymous reviewer for urging us to clarify this point.
[6] See *The International Committee of the Red Cross Database on the 161 Rules and Practice of Customary International Humanitarian Law*: https://ihl-databases.icrc.org/customary-ihl/eng/docs/home.
[7] Moreover, those designated entities who provide aid in armed conflict will not be harmed. For example, the emblem of the "Red Cross," or of the "Red Crescent," shall be required to be respected as the sign of protection.

mental torture, corporal punishment, or cruel or degrading treatment.

6. Parties to a conflict and members of their armed forces do not have an unlimited choice of methods and means of warfare. It is prohibited to employ weapons or methods of warfare of a nature to cause unnecessary losses or excessive suffering.
7. Parties to a conflict shall at all times distinguish between the civilian population and combatants in order to spare civilian population and property. Neither the civilian population as such nor civilian persons shall be the object of attack. Attacks shall be directed solely against military objectives (de Preux 2023, 7).

The principles we derived from JWT, coupled with further insights from IHL, provide utilitarian (ends-based), deontological (duty-based), and prudential criteria for determining whether an exception to selling arms can be made. In what follows, we elaborate on how these considerations can be translated to specific decision criteria for customer acquisitions and usage policies.

2. A Framework for Responsible Customer Acquisitions (Component 1)

2.1 Criteria

To begin, it must be noted that knowingly selling a product to a morally compromised buyer can make the seller complicit in that buyer's wrongdoing. This is most obvious when the product itself has the capacity to inflict harm.[8] But it can be true even if the product itself has no harmful capability. Consider that selling office supplies to a drug cartel will facilitate the cartel's criminal projects despite the limited direct danger of the product itself. While a seller might minimize potential complicity by securing the product itself against misuse, to the extent that this is feasible, protecting against misuse also requires a customer's willing compliance. It requires that customers can be trusted to use products safely and for rightful ends.

How reliable must a customer be, and how can we predict this? Even customers with perfect track records have the potential for human error and technical failure. Moreover, there is hardly any entity among potential consumers of weapons technology that has not engaged in wrongdoing at some point in its history. Sweden, for instance, is often touted as a "humanitarian superpower" for its strong commitment to human rights both domestically and globally (Simons and Manoilo 2019, 1-3). While this is certainly true in many ways, even this apparently "easy" case is more complex than it may first appear. For example, Sweden participated in the slave trade during the eighteenth and nineteenth centuries (McEachrane 2018) and has been criticized for cooperating with the US's torture of terrorism suspects in the early 2000s (Human Rights Watch 2006). The challenge, therefore, is to establish a reasonable threshold for a customer's reliability.

We thus sought to identify specific criteria that can ground a belief in the trustworthiness of a potential customer. Striving to balance theoretical rigor with practical feasibility, we considered not only how potential elements might relate theoretically to the concept of trustworthiness, but also how potential elements might be measurable empirically. While an organization can never have complete certainty over how its product will be used by its customers, the aim of this procedure—when combined with components governing usage, monitoring, and enforcement—is to provide reliable assurance of responsible use on the basis of precise theoretical standards and objective metrics. As elaborated below, the proposed criteria included: a prospective customer's baseline ethical commitments, human rights record, governance record, and military conduct record.

Importantly, because the criteria we proposed were derived from reports on national governments, we assumed that customers would mainly be units of national militaries. Absent robust metrics on the track-records of other customer types, we cautioned severely against considering customers other than national militaries.

[8] For a detailed discussion on individual complicity in collective action, see Kutz (2000).

2.1.1 Baseline Commitments

Baseline Commitments refers to whether a prospective customer operates in a jurisdiction that has committed to relevant international covenants on human rights and IHL. Major human rights treaties enjoy broad support from diverse societies and experts. They codify basic and widely accepted standards for the protection of human life and dignity. A state or subnational agent that disavows a major human rights treaty invites initial doubt about its genuine commitment to human rights and its intentions to use military technology responsibly.

There are many international human rights covenants, varying in their subject matter and global acceptance. An excellent example is the International Covenant on Civil and Political Rights (ICCPR). The ICCPR establishes minimal standards of respect for human life, liberty, dignity, and equality, rights to national self-determination, and prohibitions on torture and genocide (United Nations 1966). However, using the ICCPR as a benchmark for weighing responsible international actors in this context is ineffective, as the ICCPR has near-universal membership with only some notable exceptions. This gives the metric little means to distinguish between types of actors and offers only weak assurance that a given member state's actual commitments are reflective of its historical record or present realities.

We thus proposed adopting a more demanding measurement of a customer's baseline commitment to responsible use—particularly as it directly relates to the transnational sale of weapons—in the form of the Arms Trade Treaty (ATT). The ATT represents a commitment to limit arms sales to guard against human rights violations, war crimes, crimes against humanity, and peace and security more generally (United Nations 2013). The Treaty reaffirms commitments to international human rights declarations and announces further commitments to align arms sales with these declarations. At the time of our proposal, the ATT had been ratified by 106 of 194 states.[9] The Treaty enjoys the support of all advanced democracies except the United States, which notably and recently withdrew its signature in 2019 at the instruction of former President Trump.

Beyond the threshold of commitment to responsible use indicated by ratification of the ATT, additional signs of commitment include the Additional Protocol II of the Geneva Conventions relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) and, in particular, the Rome Statute of the International Criminal Court. Ratification of Protocol II indicates a commitment to extending international humanitarian law domestically in the treatment of a state's own residents (ICRC Database 1977). Ratification of the Rome Statute indicates a commitment to accountability for war crimes and crimes against humanity (United Nations 1998).

Using the ATT as a standing metric for a nation's baseline commitments, combined with the additional ratification of these further statutes, delivers an effective range of measurement to anchor this variable of the framework.

2.1.2 Human Rights Record

It is one thing to commit publicly to human rights and another thing to demonstrate fidelity to that commitment. A case in point is North Korea, which ratified the ICCPR in the early 1980s but has one of the most egregious human rights records in modern history (United Nations 1966). The stronger a prospective customer's record of respecting and realizing human rights, the more trustworthy that customer is as a user of dangerous technology.

What does it mean to have a strong human rights record? As mentioned above, Sweden is widely recognized as a leader in this area, but that is a relatively recent development. Other states,

[9] Given that this paper lays out the decision framework as it was proposed to our client in 2020, the data we reference in the body of this paper is the data that was current at the time of the proposal. Where appropriate, we include references to updated reports and statistics.

such as Hungary and Turkey, demonstrated a stronger commitment to human rights in the past than they have recently. Still others, such as South Africa and Colombia, may have once had a mediocre record but show strong signs of recent progress. We thus proposed that assessing a prospective customer's human rights record should account for an amalgam of (1) its historical record, (2) its current record, and (3) its projected future record or current trendlines.

There have been many attempts to track international progress on human rights. The most comprehensive current system is the Human Rights Scores data set (Fariss 2019a).[10,11] This scoring system combines many indicators from different data sets on human rights to construct a composite score for each country since 1946. The variables include extra-judicial killings, one-sided state killings, political disappearances, torture, imprisonment, massive repressive events, genocide, and politicide, among others. The index ranges from -5 to +5. Any state in the positive range can be considered to demonstrate commitment to upholding human rights, while any state in the negative range can be considered to demonstrate worsening failures to respect human rights.

2.1.3 Governance Record

Other things equal, a well-governed state is more likely to use defensive technology responsibly than a poorly governed state. Well-governed states reliably make and enforce decisions using procedures that are transparent, accountable, and fair (North, Wallis, and Weingast, 2009; Przeworski, Stokes, and Manin, 1999). This includes the rule of law and rotation of political leadership through periodic elections—the basic elements of constitutional democracy (Dahl 1998). Well-governed states subject military and police conduct to laws promulgated by elected representatives, overseen by an independent judiciary, and scrutinized by a free press and civil society (Huntington, 1957). Although alternative regimes—such as dictatorships and illiberal democracies—are sometimes capable of achieving impressive outcomes, their success is inherently unreliable (Levitsky and Way, 2010) and their commitment to human rights is compromised by definition (Sen, 1999). Constitutional democracies are far from perfect. But they are both more theoretically consistent with human rights and historically more successful at protecting them (Christiano 2011).

As with assessments of human rights records, we believe that assessments of governance should take account of historical records, current records, and projected future trends. A current-year score alone gives no indication of whether that score is an aberration from the past or likely to change.

As a barometer of good governance, we proposed using Freedom House's Freedom in the World Index, a composite of several indicators of good governance (Repucci 2020).[12] These include measurements of corruption, free and fair elections, freedom of press and expression, the rule of law, civil liberties, and political equality, among others. The Index's total scores range from 0 to 100, with 0 representing countries with no civil and political freedom whatsoever, and 100 representing near-perfect records of civil and political freedom. A score of 70 or above means that a state is considered "free" overall, with scores below 70 corresponding to "partly free" or worse.

2.1.4 Military Conduct Record

Even among states with strong human rights and governance records, military conduct varies—at times significantly. Militaries consist of large, complex bodies, with varying degrees of autonomy from civilian authorities, and varying histories and internal controls resulting in pointed

[10] For elaboration, see Fariss (2019b). For visualization of the data, see Global Change Data Lab (2022).
[11] Since 2020, a more recent data set has been made available. See Fariss et al. (2020b) as well as Fariss et al. (2020a) for elaboration.
[12] See Repucci and Slipowitz (2022) for a more recent report.

conduct variance (Huntington 1957; Feaver 1999). Most militaries run themselves in large part according to their own codes of conduct, which have evolved over many years (Janowitz 1960). Some militaries adhere principally to the rule of law, while others are more loyal to strong personalities (Desch 1999). Militaries also consist of multiple branches with different functions and professional standards, resulting in variation in conduct even within the same institution (Avant 1994). Even within militaries that uphold high standards of ethical conduct, rogue units and breakdowns in discipline are not uncommon (Arkin 2009). Level of experience and professionalism may also vary (Burk 2002).

Presently, there is no global index for rating the conduct of national militaries. However, the Geneva Centre for the Democratic Control of Armed Forces (DCAF) is a Swiss organization that monitors militaries and publishes its findings in periodic reports.[13] These reports are available online and can be reviewed for guidance on particular militaries of interest. Other INGOs, such as Human Rights Watch, provide annual reports on each country's overall human rights record, which will usually cover noteworthy conduct by national militaries and police. Our recommendation was that the service provider utilize a combination of such reports, where available, to determine a prospective customer's military conduct record.

2.2 Proposed Customer Screening Procedure

How should the proposed assessment criteria be aggregated to support sound conclusions? Averaging across the criteria would be a mistake, as this would allow for the possibility that prospective customers' achievements in some criteria would outweigh egregious failures in others. A more promising option is to establish a threshold or red line for each criterion. Any prospective customer that falls below this threshold on any criterion represents an untrustworthy partner and should be disqualified. This approach need not assume that the criteria are of equal weight or that they are perfectly commensurable. It simply assumes that there is some point on each criterion that cannot be crossed without indicating serious cause for alarm. There is reasonable room for debate on where exactly to set the thresholds. We proposed the following.

2.2.1 Red Line Test for Initial Consideration

To meet the *Baseline Commitments threshold*, a prospective customer must have ratified the ATT to be above the red line. Ratification of identified additional statutes can further increase confidence in a customer's baseline commitments.

To meet the *Human Rights Record threshold*, a prospective customer must achieve a 0 or higher on a composite index of Human Rights Scores (aggregating separate scores for historical data, current score, and current trajectory). To create this composite, one can examine the average of the country's scores since 1990, its current score, and the rate of change since 2010.

To meet the *Governance Record threshold*, a prospective customer must achieve a 70 or higher on a composite index of Freedom in the World Scores (aggregating separate scores for historical data, recent data, and current trajectory). Freedom in the World data is available since 2006.

To meet the *Military Conduct Record threshold*, a prospective customer must have no record of systemic failures of military conduct in its recent history, as indicated by reports from DCAF or similar expert bodies and related reports.

Prospective customers that pass all these gates can then be considered eligible for further consideration (Table 1). Requiring that all four thresholds be met for consideration provides a chain or sequence of robust backstops to assure the service provider that its products will not be misused.

[13] See http://dcaf.ch for more information.

Table 1. Thresholds for Initial Consideration of a Potential Customer

CRITERION	NAME	KEY INDICATOR	THRESHOLD
A	Baseline Commitments	ATT	Ratification
B	Human Rights	Human Rights Scores	Composite of historical, current, and projected score above 0
C	Governance	Freedom in the World Index Total Score	Composite of historical, current, and projected score 70+
D	Military Conduct	DCAF, HRW, and similar expert reports	No recent reports of systemic misconduct

2.2.2 Graduated Assessment above the Thresholds

Prospective customers that meet the threshold of each criterion can also be compared on a graduated scale (Table 2). Those with especially robust records on human rights, governance, and military conduct will naturally merit greater consideration than those who barely meet the thresholds for these criteria.

Beyond ratification of the ATT, prospective customers merit additional consideration if they have also ratified Protocol II, the Rome Statute, or both.

For states that operate in the positive range of the Human Rights Scores index, the median score in 2017 was 1.4. We proposed this score be used as a benchmark to assess the strength of prospective customers' human rights records. Those with composite scores above 1.4 can be considered especially strong on this indicator.

Table 2. Scale for Comparing Customers that Exceed the Thresholds

CRITERION	NAME	KEY INDICATOR	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments	ATT, Protocol II, Rome Statute	Non-ratification of ATT	Ratification of ATT alone	Ratification of the ATT and one or more other treaty
B	Human Rights	Human Rights Scores	Composite of historical, current, and projected score above 0	Composite of historical, current, and projected score below 1.4	Composite of historical, current, and projected score above 1.4
C	Governance	Freedom in the World Index Total Score	Composite of historical, current, and projected score below 70	Composite of historical, current, and projected score below 90	Composite of historical, current, and projected score above 90
D	Military Conduct	DCAF, HRW, and similar expert reports	Recent reports of systemic misconduct	Reports of isolated misconduct	No recent reports of misconduct

For states that operate above 70 on the Freedom in the World index, the median score in 2020 was 90. We proposed this score be used as a benchmark to assess the strength of prospective customers' governance records. Those with composite scores above 90 can be considered especially strong on this indicator.

While the threshold for military conduct is the absence of recent reports of systemic misconduct, prospective customers that pass this test can be assessed against whether they have isolated reports of misconduct. Customers with strong records on this count can be considered especially responsible in this area. However, it will be worth taking account of the size of a customer's military when making these assessments. The larger a military, the higher the likelihood that someone will have committed some kind of wrongdoing—even if overall standard of conduct is excellent.

2.3 Examples for Testing and Illustration

To test this methodology, we first offered four sample cases, using countries with mixed track records in governance and human rights that otherwise fit the profile of potential customers for our client. We then offered a fifth case to have a standard control to compare against—a prominent NATO country that would likely be considered a reliable partner, before such an exercise. After working through how the proposed method would judge each case, we analyzed how well these conclusions match our own considered judgments.

2.3.1 Test Case: Qatar

Qatar has not signed or ratified the Arms Trade Treaty (United Nations 2013). This placed Qatar below the red line for Criterion A.

Qatar's Human Rights Score (ranging from -5 to +5) has been positive since its independence in 1971. It has hovered around 1.5 since 1990. Recent trends have been inconsistent, dipping from 1.9 in 2010 to several years at 1.3, and rising to 1.6 in 2017 (Schnakenberg and Fariss 2014; Fariss 2019a).[14] But this report did not indicate any impending decline below 1.0. Furthermore, Qatar's Composite Human Right's Score is 1.56, well above the threshold of 0.[15] We thus determined that Qatar could be considered a consistent respecter of human rights overall, placing the country above the red line for Criterion B.

Qatar's Freedom in the World Score (ranging from 0 to 100) has ranged between 25 and 28 since 2006, the first year of available data (Freedom House n.d.a). Qatar's Composite Freedom in the World Score is 26.13, well below the minimum threshold of 70.[16] Qatar clearly counted as "not free," and did not meet the threshold for Criterion C.

While DCAF did not have any public reports with information on Qatar's military conduct at the time of our proposal, a 2020 report by the U.S. Congressional Research Service (an independent and nonpartisan body) indicated that Qatar is the second-largest recipient of U.S. arm sales and frequently receives training from U.S. personnel (Katzman 2020). Although the Qatari military's

[14] At the time of the proposal, the most recent year for which data was available to compute Human Rights Scores was 2017. See Global Change Data Lab (2022) for more recent scores.

[15] To generate the Composite Human Rights Scores, we averaged the country's mean score from 1990-2017 (to reflect its historical record), its 2017 score (to reflect its current record), and the country's mean score from 2010-2017 (to reflect its projected record).

[16] To generate the Composite Freedom in the World Scores, we averaged the country's mean score from 2006-2020 (to reflect its historical record), its 2020 score (to reflect its current record), and the country's mean score from 2013-2020 (to reflect its projected record).

conduct record is generally unremarkable, Qatar is part of the Saudi-backed military coalition against the Houthi rebels in Yemen, a coalition that has been accused of war crimes (Human Rights Watch 2018). That relationship gave pause; but overall, their military ethical use record was above the minimal threshold for Criterion D and approached a strong rating.

Table 3. Indicators of State Customer Trustworthiness: Qatar

CRITERION	NAME	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments	⊗		
B	Human Rights		✓	
C	Governance	⊗		
D	Military Conduct		✓	

Applying the red line test to Qatar suggested that Qatar was not, at the time of our proposal, a trustworthy customer to partner with (Table 3). It had not demonstrated a baseline commitment to responsible arms trade and its record on governance remained exceptionally poor. We thus strongly recommended against doing business with Qatar until a reassessment is appropriate.

This conclusion made sense in light of the broader context. The country's ongoing mistreatment of noncitizens and ethnic minorities (United Nations 2020), coupled with a lack of protective mechanisms—such as electoral accountability, an independent judiciary, or a free press—inspired insufficient trust in Qatar's reliability (Freedom House n.d.b).

2.3.2 Test Case: Colombia

Colombia has not ratified the Arms Trade Treaty (United Nations 2013). This placed Columbia below the red line for Criterion A.

Colombia has an historic Human Rights Score hovering around -2. However, this does not paint the full picture, as, at the time of analysis, Colombia's score had risen consistently every year since 2003, reaching -0.89 in 2017 (Schnakenberg and Fariss 2014; Fariss 2019a). Averaging the historic, current, and projected records resulted in a Composite Human Rights Score of -1.4. While still below the red line for Criterion B, Columbia demonstrated movement in a positive direction.

Colombia's 2020 Freedom in the World Score was 66, qualifying it as "partly free." While its historical average is 62.5, its score has increased steadily since 2013, when it stood at 61 (Freedom House n.d.a). Averaging the historic, current, and projected records resulted in a Composite Freedom in the World Score of 64, which placed Columbia below the red line for Criterion C. However, at the time of analysis, we predicted that if progress continued at the same average rate since 2003, it would cross the threshold of 70 by 2026.

While DCAF materials did not offer a comprehensive picture of Colombia's military conduct, Human Rights Watch reports that wanton killing of civilians by army brigades was a common practice between 2002 and 2008 (Vivanco 2019). Since the government's historic peace accord with the FARC militia in 2016, the country's security forces have shown restraint even as militia members have sought to resume hostilities (Human Rights Watch 2020b). Meanwhile, the National Police have been accused of abusing peaceful demonstrators with impunity (Human Rights Watch 2021). We determined that Colombia met the threshold for Criterion D; however, the ongoing

political instability indicated that there was still significant room for improvement in this area.

Applying the red line test to Colombia suggested that Colombia was not a trustworthy partner, as it remained below the thresholds on Human Rights and Governance and has yet to ratify the Arms Trade Treaty (Table 4). Encouraging trends suggested that Colombia could become a trustworthy partner in the near future, however.

Table 4. Indicators of State Customer Trustworthiness: Colombia

CRITERION	NAME	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments	⊗		
B	Human Rights	⊗		
C	Governance	⊗		
D	Military Conduct		✓	

This conclusion made sense in light of the broader context. Although Colombia has shown consistent progress in overcoming its deeply checkered past, it remains a conflict-ridden country with numerous areas of instability, and relapses are not unlikely. With further progress, though, Colombia could very well achieve greater stability and merit reconsideration as a potential customer.

2.3.3 Test Case: Peru

Peru has ratified the Arms Trade Treaty (United Nations 2013), Protocol II (ICRC Database 1977), and the Rome Statute (United Nations 1998). This placed it in the "Excellent" category for Criterion A.

Peru has an historical Human Rights Score of -0.58 and a 2017 score of 0.74. Its score has risen consistently since 1990, crossing the 0 threshold in 2005 (Schnakenberg and Fariss 2014; Fariss 2019a). Averaging the historic, current, and projected trends resulted in a Composite Human Rights Score of 0.16. There have been some difficult years since 2005, but Peru never quite fell below 0 again, and, at the time of analysis, had returned to its earlier highs and beyond most recently. We thus concluded that Peru's Human Rights Record could be considered acceptable and rising, placing it above the threshold for Criterion B.

Peru's 2020 Freedom in the World Score was 72, which is also its historical average (Freedom House n.d.a). Recent trends at the time of analysis suggested no clear movement in either direction on governance. Peru qualified in the "free" category on this metric and has had remarkable stability, though there remains significant room for growth. Peru's Composite Freedom in the World Score is 71.96, which placed it near the bottom of the acceptable range for Criterion C.

While DCAF materials did not offer a comprehensive picture of Peru's military conduct, a 2020 Human Rights Watch report determined that Peru continues to reckon with its conflict-ridden past. Hostilities ceased in 2000, but Peru has made little progress toward holding perpetrators accountable for war crimes and crimes against humanity. At the time of our proposal, reports indicated that military and police conduct since 2000 had generally been responsible, with certain exceptions. Police had been accused of excessive force and extra-judicial killings, but reports were

both sporadic and declining (Human Rights Watch 2020d).[17] We determined that Peru met the threshold for Criterion D but still had room for improvement.

Table 5. Indicators of State Customer Trustworthiness: Peru

CRITERION	NAME	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments			✓
B	Human Rights		✓	
C	Governance		✓	
D	Military Conduct		✓	

Applying the red line test to Peru suggested that Peru was a potentially trustworthy customer to partner with (Table 5). This conclusion made sense in light of the broader context. Peru's poor historical Human Rights Score is due to its civil war, which concluded in 2000 (Human Rights Watch 2020d). Its record since has been encouraging. We thus recommended that a partnership with Peru should be approached with caution, and that relevant restrictions should be applied in the course of the relationship, as described further below.

2.3.4 Test Case: Chile

Chile has ratified the Arms Trade Treaty (United Nations 2013), Protocol II (ICRC Database 1977), and the Rome Statute (United Nations 1998). This placed Chile in the "Excellent" category for Criterion A.

Chile has an historical Human Rights Score of 0.6 and a 2017 Human Rights Score of 1.1 (Schnakenberg and Fariss 2014; Fariss 2019a). Chile has had durable stability and a strong regional-relative score for more than two decades. While there is still substantive room for growth, Chile's Composite Human Rights Score of 0.89 placed it well above the red line threshold. We thus considered it to be at the upper end of the acceptable range for Criterion B.

Chile's 2020 Freedom in the World was 90, and its historical average is 95. However, after several years of slight decline, Chile tumbled from 94 to 90 in 2020 (Freedom House n.d.a). This is reflected in Chile's Composite Freedom in the World Score, which is 93. Even with the possibility of further decline in the future, we determined that Chile should be placed in the "Excellent" category for Criterion C.

While DCAF materials did not offer a comprehensive picture of Chile's military conduct, Human Rights Watch reported that Chile continues to reckon with its conflict-ridden past and has made little progress toward holding perpetrators accountable for war crimes and crimes against humanity (Human Rights Watch 2020a). Some worry that the military has yet to fully embrace democratic institutions following the collapse of the Pinochet regime (Feinberg 2019). Military and police conduct in recent years has generally been responsible, with the exception of reports of excessive force by police against demonstrators and detainees (Human Rights Watch 2020a). We thus determined that, while room for improvement remained, Chile met the threshold for Criterion D.

Applying the red line test to Chile suggested that Chile was potentially trustworthy, scoring

[17] Since 2020, human rights violations and reports of excessive force by police in Peru have risen. See Human Rights Watch (2023) for more detail.

above the threshold in all four categories and in the "Excellent" range in two (Table 6). This conclusion made sense in light of the broader context. After casting off its military dictatorship in 1990, Chile has become the leading Latin American country on numerous indicators of peace, progress, and prosperity. However, although Chile showed no sign of falling below the red-line threshold on Governance, we advised that its recent decline should be monitored for further developments.

Table 6. Indicators of State Customer Trustworthiness: Chile

CRITERION	NAME	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments			✓
B	Human Rights		✓	
C	Governance			✓
D	Military Conduct		✓	

2.3.5 Test Case: France

France has ratified the Arms Trade Treaty (United Nations 2013), Protocol II (ICRC Database 1977), and the Rome Statute (United Nations 1998). This placed France in the "Excellent" category for Criterion A.

France has an historical average Human Rights Score of 1.3 and a 2017 score of 1.8. Its score varied considerably throughout the 1990s, reaching a low of 0.8 in 2003 (Schnakenberg and Fariss 2014; Fariss 2019a). France has a Composite Human Rights Score of 1.6, and its score has grown consistently over the past decade. While there is still room for growth, France was placed in the "Excellent" category for Criterion B.

France's 2020 Freedom in the World Score was 90, while its historical average Freedom in the World Score is 93. France had been in slow decline from 96 in 2010 to 90 in 2020 (Freedom House n.d.a). This is reflected in its Composite Freedom in the World Score, which is 92. Given this downward trend, we recommended that France could be considered at the high end of the "Mid-Range" category or the low end of the "Excellent" category for Criterion C.

While DCAF materials did not offer a comprehensive picture of France's military conduct, Human Rights Watch reported no concerns with France's military. There were, however, several concerns with France's police, including mistreatment of refugees, racial discrimination, excessive force (Human Rights Watch, 2020c), and abuse of counterterrorism provisions—encouraged by a 2017 statute that expands police powers and protects them from oversight (United Nations 2019). We thus concluded that France could be considered in the "Excellent" range for Criterion D, but recommended caution if paramilitary or police forces were considered.

Table 7. Indicators of State Customer Trustworthiness: France

CRITERION	NAME	RED LINE	MID-RANGE	EXCELLENT
A	Baseline Commitments			✓

B	Human Rights			✓
C	Governance			✓
D	Military Conduct			✓

Applying the red line test to France suggested that France was likely trustworthy, scoring in the high end of the range for Baseline Commitments, Human Rights, and Military Conduct (Table 7). Given its strong performance on other areas, we determined that its recent decline on Governance may not be cause for significant alarm. This conclusion made sense in light of the broader context. Despite its colonial past and ongoing challenges with social inclusion, France has long been at the forefront of arms control, human rights, democratic governance, and military ethics.

2.4 Trustworthiness Tiers

As the foregoing examples demonstrate, countries that exceed the thresholds for trustworthiness show considerable variation in their achievements across the proposed criteria. Countries that score highly on several criteria arguably merit greater initial trust than countries that score poorly on most criteria. High-achieving countries might be treated as stronger contenders for business or—as we suggest later—eligible for slightly more lenient tracking and reporting requirements. Performance among above-threshold potential customers can be broken down into the following two tiers.

Tier I: Standard. Prospective customers that score "Excellent" on three or more of the criteria for trustworthiness should be considered ideal partners for the service provider. This metric was chosen as a benchmark as only a handful of potential customers will score "Excellent" in all four categories. Failure to achieve excellence in a single category can sometimes be attributed to measurement error. It can also be excusable for other reasons. For instance, countries occasionally refuse to sign treaties on the basis of technical or political grounds, rather than out of substantive disagreement (e.g., Lake 2001; Simmons 2009).

Tier II: Cautious. While failure to achieve excellence on a single criterion might be excusable, underperformance on two or more categories begins to look like a pattern. We proposed that countries that receive an "Excellent" rating on two or fewer criteria be treated with more caution.

We proposed that regular reinvestigation and assessment of ongoing customers should be conducted and noted the possibility that customers could move or change status from one tier to another depending on new information and behavior trends.

3. A Framework for Responsible Usage Policies (Component 2)

Once the service provider has determined which entities merit consideration for a product sale in accordance with Component 1, a next step is to establish guidelines about how customers may use the product. The rationale behind this step is that a seller's interest in avoiding complicity in the potential wrongdoing of prospective customers can override a buyer's interest in using a product however they see fit. This position is not uncontroversial, as buyers do have legitimate interests in purchasing products freely and equitably. However, we cannot overlook the fact that these buyers have the potential to cause great harm and violate the rights of third-party innocents. And this consideration merits special attention.[18]

[18] Kutz (2000) offers a relevant case by way of example, in which a gun merchant sells a gun to someone who proceeds to carry out a robbery. While he acknowledges that the gun merchant may not be fully

Strong scrutiny at the first stage should make further stages significantly easier. However, grounds for caution remain. Even a customer that scores highly on Component 1 and displays strong motivation for responsible use may find this aim difficult to achieve without clear guidelines for responsible use. We thus proposed the following Acceptable Use Policy (AUP), drawing once again on JWT and IHL.

3.1 Proposed Elements of an Acceptable Use Policy

First, the customer must agree not to use the products for any purpose other than legitimate military and defense purposes, in accordance with (1) International Law on the use of force in international relations and (2) all relevant Laws of Armed Conflict and statutes. These include: The Geneva Conventions (First, Second, Third, and Fourth; and Additional Protocols I, II, and III), as well as all "general practices accepted as law" known as Customary International Humanitarian Law as outlined by the International Committee of the Red Cross Customary Law and Practice Database.

The customer must also ensure that any use is only ever deliberately directed towards combatants and is never intentionally directed against those non-combatant soldiers or others, such as civilians, aid workers, or journalists, who should be provided with protected status. Only those who through their actions and/or membership in a hostile party may be targeted, and only so far and for as long as they represent a genuine threat.

The customer must agree that all operators of any product are certified in responsible product use. They must also be able to regularly verify the following training for each operator on an ordered and agreed upon schedule: training on the functionality, safety, and proper operation of the product; training on the Law of Armed Conflict, all relevant elements of Customary International Humanitarian Law, and all theatre-specific Rules of Engagement; and training on the responsible use of autonomous weapons and dual-use technology.

The customer must agree to monitoring and reporting requirements in accordance with Component 3 (see §4.3). Finally, the customer must agree to penalties and remedies for findings of noncompliance in accordance with Component 4 (see §5.2).

3.2 Proposed Levels of AUP Stringency Agreements for Different Customer Tiers

To give greater assurance and better guard against misuse, we proposed more stringent agreements, policies, and requirements for customers in Tier II than those required for customers in Tier I (more on this below). These agreements with a given customer could be re-evaluated as deemed appropriate depending on how successful the customer is at demonstrating trustworthiness in accordance with Component 1 metrics.

4. A Framework for Monitoring Product Usage (Component 3)

Once a service provider has established guidelines on how customers may use its products, a next step is to consider whether and how the provider should verify compliance with these guidelines. In the presence of regulatory voids, service providers have a moral duty to ensure their products are being used responsibly, especially if the provider has reason to suspect that a particular customer may misuse the technology. Failure to act in the face of such suspicions will make the

---

complicit in the crime—at least not to the extent the robber is—he nevertheless remarks that there is "something to the idea that the gun seller is morally complicity in the crime—indeed, intuitions are strong when injuries or death result" (169). Later, he adds that "[a] gun seller's refusal to associate himself with even inevitable crime identifies him with the interest of those who would be harmed" (190). We agree with Kutz that the mere fact that service providers may not themselves use weapons technology for harmful ends does not absolve them of all responsibility for how their product is used, and we maintain that it should be a priority of the service provider to align their interests with those of vulnerable parties.

provider morally complicit in any resulting harm. In order to prevent or limit acts of misuse, the provider will need to gather reliable information on how customers are in fact using its products.

One way to begin thinking about procedures for tracking usage is to identify ideal epistemic standards: the information the provider would ideally need to form the best judgments regarding customer compliance. Having articulated these ideals, we can then adjust them to incorporate countervailing values and technical limitations.

4.1 Theoretical Considerations

4.1.1 Ideal Information to be Collected

What information the service provider would need to determine compliance with usage guidelines would depend on precisely how Component 2 is defined and reflected in a customer's contract. Presumably, however, the service provider would want data on the following questions:

- Has usage resulted in harm to any persons or property? If so, has this harm been consistent with all relevant laws and principles of armed conflict?
- Who has used the product?[19] Has the product been shared with third parties? Have all users and operators been duly authorized and certified according to Component 2 guidelines and required trainings? Have all uses occurred with proper oversight or chain-of-command in accordance with Component 2 guidelines?
- Has the product performed as intended? Under what circumstances have technical failures or malfunctions occurred? Could technical changes reduce future error or unintended harm?

4.1.2 Customers' Interests in Privacy

Once the service provider has determined the information it would ideally need, a next step is to consider whether collection of this information could interfere with a customer's legitimate interests—for example, protecting information that could unduly compromise national security, the integrity of its missions, or the safety of its personnel. However, not all claims to privacy are justifiable, and customers' legitimate interests do not preclude the sharing of certain information with the service provider or other external monitors. The costs to the customer of some forms of legitimate information sharing may be outweighed by the interests in ensuring compliance with the laws and principles of armed conflict, especially when there are reasons to suspect noncompliance with key requirements.

Certainly, the potential for conflict will be limited if the service provider can make reasonable assurances of confidentiality, either in the way it handles sensitive information or in the kinds of information it collects. For instance, the service provider should put in place robust protocols for protecting secure access to customer information and provide assurance of this security to its customers. The service provider may also seek to acquire only data that is in anonymous or summary form, so as not to be able to identify specific users, targets, or casualties. Relatedly, it could feed raw data into a statistical model that returns probabilities of potential misuse for any given incident without any identifying sensitive information about that incident.

4.1.3 Technical Feasibility

A third set of considerations involve what forms of information collection are technically

[19] Although this question might seem redundant in light of the previous one, which would assess (e.g.) whether force had been undertaken by a "legitimate authority" as required by the laws and principles of armed conflict, in fact the question seeks to determine whether such an authority has delegated its powers responsibly. We thank an anonymous reviewer for pressing us on this point.

feasible, in the near or future term iterations of products, and the costs associated with each possibility. Suppose that the service provider could monitor each use of the product in real time, using analytic technology to determine the identities of each user, target, casualty, and so forth, and produce near real-time incident reports. Even if this were both consistent with customers' legitimate interests and technically feasible, it might require making the product less agile or too expensive. So, it is imperative to consider technical possibility along with the corresponding trade-offs to be made in terms of costs and other product features for any technical tracking and monitoring options.

4.2 Reporting and Tracking Options

Several strategies for data collection exist. We outlined the following as viable options for the service provider.

4.2.1 Customer-Provided Reports to the Service Provider

Customer-provided reports are the least desirable option for verifying compliance, as customers will have incentives to omit, delay, or alter the reporting of unflattering information. However, the prospect of providing reports may have a modest effect on accountability, as customers may be less inclined to abuse the technology if they know they must report their usage in the future.

4.2.2 External Audits of Customers by a Third Party

External audits by qualified third parties could mitigate some of the foregoing problems. Auditors might of course request customer-generated reports. However, auditors might also be empowered with more expansive powers to review customer documents and procedures in addition to any customer-provided reports. The status of an auditor as a trusted impartial expert could make customers more willing to submit to scrutiny (though such an audit may also raise confidentiality and data security concerns raised above). By sharing the burden of judgment with a third-party expert, the service provider could also limit its own potential complicity.

4.2.3 Collection of Data on a Chip that is Shared as Needed

Data could also be collected on a natively stored chip that is shared with the service provider (or a third-party auditor). There are choices to be made about what kinds of data are stored and the conditions under which these data are shared. Data collection could be expansive or limited. Expansive data collection might involve audio-visual recordings that the operators see, as well as user statistics, target statistics, and firing statistics. Data collection could also be limited to such things as whether users were duly authorized and whether firing rates were consistent with expectations about proportionality. Data collection could also be selective or shared with the service provider only under certain circumstances—such as in response to particular incidents, accusations of misuse, or suspected omissions in customer-generated reports. Finally, some form of permanent usage records that could be imprinted into data storage on the products themselves could be utilized (similar to "black box" recording systems on aircraft). This information could then be recovered after incidents if the product hardware itself can be retrieved.

4.2.4 Collection of Data that is Shared Wirelessly in Real Time

Real-time (or near real-time) data sharing with the service provider (or a third-party auditor) would provide the best measure of a customer's compliance—especially if that data is comprehensive. Customers could have valid objections to comprehensive sharing of data in real-time. If there are risks of data leakage, this could reveal targets or tactics to third parties. This would be a significant potential security risk. But there may also be ways around these problems, say, by collecting only anonymized or summary data, or by subjecting transmission to a time-delay. The precise kinds and amounts of data to be transmitted in real-time may also be limited by the

availability of transmission services, on-board data storage, or electricity. Both this monitoring option and the previous, while recommended, deserve further technical feasibility confirmation and exploration.

4.3 Proposed Monitoring and Tracking Procedures

None of the foregoing options are mutually exclusive, and the service provider could profitably rely on some combination of all of them. We ultimately proposed a baseline policy composed of the following six elements. First, the customer should provide annual reports on how the product has been used, including the missions in which it was deployed, the justification for the use of the product in each instance, and the casualty counts and distributions, and damage to property.

Second, the customer should file an incident report with the service provider of any incident that results in harm to persons and (possibly) damage to property above a certain damage threshold. Third, reports must be accompanied by sharing access to the product's internal chip, which stores detailed data on how the product has been used, including the distance of targets, information used to select targets (infrared data, images, audio, etc.), and information used to verify target hits (infrared data, images, audio, etc.).

Fourth, the service provider should employ an independent auditor to verify customer reports and chip data. This should occur both on a random basis for all customers in Tier II (Cautious), and whenever there are grounds for suspicion of gross misuse or noncompliance. Fifth, in order for the product to be operated, users must be verified by logging into a server controlled by the service provider where operator training certification is tracked and monitored, in accordance with the AUP outlined in Component 2. Sixth, the service provider should engage in wireless tracking of which authorized users are logged in, the geographic location of the product when in use, and rounds fired.

5. A Framework for Enforcing Compliance (Component 4)

5.1 Background

Whether and how the service provider could enforce compliance would depend to a great extent on how much knowledge it could acquire about product usage. Without the ability to authenticate users, the service provider could not verify whether users of its products are duly authorized. Without the ability to gather comprehensive data about use, it could not verify whether misuse has occurred. And without the ability to gather that data quickly, the service provider would risk the chance that acts of misuse may continue or increase while waiting for information.

The hope and expectation were that the successful design and application of the first three components of this framework would dramatically limit the need for mechanisms of enforcement and response. The more customers are carefully vetted, bound by strong contracts and agreements, and closely monitored, the less likely misuse is to occur. However, one must always prepare for the possibility that misuse will occur and have measures in place to mitigate it and respond responsibly.

It is helpful to think of enforcement as having two primary dimensions: *prevention* and *recourse*. There are in fact enforcement measures a service provider can take to prevent or limit potential for acts of misuse. Chiefly, a service provider can program its products to require user authentication, as discussed further below. Beyond authenticating users and ensuring that operators have required training protocols, we believe there is little that the service provider can do to prevent imminent acts of misuse or halt such acts once in progress. However, the service provider does have several options for responding to acts of misuse after they have occurred. These responses can be helpful for preventing future misuse. Furthermore, a robust system of response protocols can have a deterrent effect on initial acts of misuse.

As this last point suggests, it would be essential that the service provider make its

enforcement policies transparent to customers. In theory, the service provider could choose not to disclose its response protocols to customers if it believed that such protocols would severely disincentivize purchases. This may be especially tempting for something like remote disabling capabilities, for example. (More on this below.) A decision to withhold this information would be ethically objectionable, however. Transparency communicates good faith, promotes trust, and permits recourses to have a deterrent effect.

The last point should not be undervalued. The technological capabilities this service provider was developing would likely be in high demand by militaries around the world. This could give the provider of such capability the high ground in not only negotiating what its users agree to, but it also strongly incentivizes customers to not lose such capabilities through misuse. This deterrent effect only works, however, if the threats of remedies for noncompliance are credible. Not only does secrecy of possible responses to misuse compromise each of these values, but it could have costly downstream effects on future customer relationships if and when the information becomes public.

5.2 Proposed Response Sequence

5.2.1 User Authentication

As a preventive measure applied to all customers, we recommended that the service provider require user authentication through a passcode, bar code, or biometric scan. Failure to authenticate would thus make the product inoperable. Although no authentication process is foolproof, such a process can reaffirm the expectation that only duly authorized users are supposed to operate the product and make it more difficult for unauthorized users to do so. Authentication processes that interact with the service provider's servers also supply the service provider with useful knowledge about usage.

5.2.2 Severance of Business Relationship

If a service provider has strong reason to suspect misuse, a natural first step is to warn the customer that failure to make amends will result in punitive consequences. These consequences may include a moratorium on future product sales and maintenance. To make these warnings effective, we proposed making periodic maintenance by the service provider central to the successful operation of the product. If a customer failed to heed the warnings
within a given time period, the service provider could deny the customer future sales and maintenance.

5.2.3 Public Disclosure and Censure

While business relationships might be resumed if the customer made the necessary corrections to remain in good standing, damage to its public reputation can be more permanent. If severance of the business relationship proved insufficiently effective, we noted that the service provider could warn the customer that its conduct will be reported to the press and human rights monitoring bodies. If the customer failed to heed these warnings with a given time period, the service provider might indeed take these measures and offer public censure, describing their disagreement with and rejection of such misuse.

5.2.4 Removal of Technical Support (Effective Disabling)

When all else fails, the service provider might find itself in a position of having exhausted the above measures and yet still not be convinced that the product was being used appropriately. To guard against misuse in these most extreme cases, only after all other measures have been tried, the service provider could next simply remove technical and software support to the product. This would likely be a part of step two of the response sequence (severance of business relationship). However, it

could be done separately. The proposal was that without proper software and technical support, the drone platform would become *effectively* unusable and inoperable.

To be clear, such an act would not be an overt disabling of the craft. Rather, this would be simply the removal of technical support from the service provider, an act fully within their legal rights at any time they deem. Such removal would make the functioning of the product increasingly difficult to the point where former customers would likely abandon its use. Ethically, this would be an act of omission rather than an act of commission, such as actively disabling the craft. The distinction is important both for the service provider's own responsibility and for contractual reasons.

5.2.5 Proactive Remote Disabling by Service Provider or Customer

Proactive, direct disabling of the product remotely may seem like a drastic step, and as noted, it is highly doubtful a military customer would agree to giving the service provider such power over the product. However, if there were reason to worry that the product could be stolen, resold, or captured in battle, remote disabling might be an important feature for protecting customers' *own* interests. We argued that remote disabling capability—for the customer to initiate themselves—should be offered as a product feature to ease customer's valid concerns over the weapon system falling into the wrong hands. The customer would essentially have a remote "kill" switch they could activate at any time on the product if and when they believed they had lost effective control over the platform. This would disable the platform and ensure that it could not be used for ill purposes nor against the customer's own troops.

6. Conclusion

To date, there has been a serious lack of available resources for companies seeking consistent and ethically grounded principles for responsible use of weapons technology. This framework seeks to address that gap by offering a coherent and principled decision-making procedure for customer screening and due diligence. While utilizing this framework cannot guarantee that weapons technology will not be misused (in practice, virtually no framework could make that guarantee), we believe it does promise to provide reliable assurance of responsible use. Our hope is that sharing this framework will spark further discussion regarding the ways in which companies designing and building new technologies at the frontier can seek to navigate a marketplace where regulations may be absent, underdeveloped, or imprecise.

This case study also has some significant limitations. First, the paper lays out four thresholds that potential customers—in this case national militaries—must meet to be considered trustworthy partners. Room for debate remains as to whether these particular four are the optimal criteria and whether they are sufficiently constraining for a product as rife for potential wrongful use as a weaponized drone platform. For other kinds of dangerous technologies, presumably, the thresholds for potential customers could be less restrictive, or more restrictive, relative to the perceived risk of misuse and extent of potential harm. Other limitations include the lack of publicly available information that were used to create the series of thresholds in this case study. For example, DCAF materials were originally cited as the benchmark to be used for rating the conduct of national militaries. However, in none of the cases discussed were DCAF materials actually available, leading us to rely on more circumstantial reports from other sources instead. This general lack of available resources for assessing the conduct of military forces is regrettable.

One particularly interesting feature of this study is both the complicated relationship it exposes between safeguards at different stages of a product lifecycle and the tension between ethics-by-design and later-stage efforts to mitigate ethical risks. Although our client initially sought advice on customer acquisitions after the product was already in a late developmental stage, we advised that a responsible customer acquisitions policy could not be divorced from questions about the

capabilities of the product itself. We recommended several potential retrofits and add-ons to the product to prevent misuse and enforce responsible use. Fortunately, the stage of product development at that time still allowed for these changes to be meaningfully contemplated. However, integrating ethical considerations much earlier in the initial product design would have likely reduced costs and enabled a more holistic and seamless integration of ethical safeguards.

For new and emerging technologies that have serious risk of misuse or are inherently dangerous, it is especially important to examine ways in which they can be responsibly developed despite persistent regulatory lags. We believe that self-regulation is no substitute for democratically authorized regulation of dangerous technology. At the same time, we recognize that developers of dangerous technologies may often have an acute understanding of their ethical risks and strong motivations to mitigate them. Additionally, initial attempts by private organizations to fill regulatory voids can sometimes provide a helpful basis for development, criticism, and improvement by regulatory authorities at subsequent stages. We hope that case studies like this one can further elucidate some of these complexities involved in parceling out responsibility in evolving efforts to safeguard emerging technologies.

REFERENCES

Aquinas, T. 2002. *Political Writings*. Edited and translated by R. W. Dyson. Cambridge: Cambridge University Press.

Arkin, R. C. 2009. *Governing Lethal Behavior in Autonomous Robots.* New York: CRC Press.

Augustine, S. 2003. *City of God*. Translated by Henry Bettenson. London: Penguin Books.

Avant, D. D. 1994. *Political Institutions and Military Change: Lessons from Peripheral Wars*. Ithaca: Cornell University Press.

Burk, J. 2002. "Theories of Democratic Civil-Military Relations." *Armed Forces & Society* 29 (1): 7-29.

Christiano, T. 2011. "An instrumental argument for a human right to democracy." *Philosophy & Public Affairs* 39 (2): 142-176.

Coglianese, C., and E. Mendelson. 2010. "Meta-Regulation and Self-Regulation." In *The Oxford Handbook of Regulation*, eds. Robert Baldwin, Martin Cave, and Martin Lodge. Oxford University Press, 145–68.

Cook, Martin. 2004. *The Moral Warrior.* Albany, NY: SUNY Press.

Dahl, R. A. 1998. *On Democracy*. New Haven: Yale University Press.

de Preux, J. 2023. *Basic Rules of the Geneva Convention and their Additional Protocols.* International Committee of the Red Cross, ref. 0365. https://www.icrc.org/en/publication/0365-basic-rules-geneva-conventions-and-their-additional-protocols

Desch, M. C. 1999. *Civilian Control of the Military: The Changing Security Environment*. Baltimore: Johns Hopkins University Press.

Fariss, C. J. 2019a. "Latent human rights protection scores version 3." *Harvard Dataverse* V1. https://doi.org/10.7910/DVN/TADPGE

Fariss, C. J. 2019b. "Yes, human rights practices are improving over time." *American Political Science Review* 113 (3): 868-81.

Fariss, C. J., M. R. Kenwick, and K. Reuning. 2020a. "Estimating one-sided-killings from a robust measurement model of human rights." *Journal of Peace Research* 57 (6): 801-814.

Fariss, C. J., M. Kenwick, and K. Reuning. 2020b. "Latent human rights protection scores version 4." *Harvard Dataverse* V2. https://doi.org/10.7910/DVN/RQ85GK

Fasterling, B., and G. Demuijnck. 2013. “Human rights in the void? Due diligence in the UN guiding principles on business and human rights.” *Journal of Business Ethics* 116 (4): 799–814.
Feaver, P. D. 1999. “Civil-Military Relations.” *Annual Review of Political Science* 2: 211-241.
Feinberg, R. E. 2019. “Order from chaos: Chileans learned the right lessons after the Pinochet era.” *The Brookings Institution*, November 18. https://www.brookings.edu/blog/order-from-chaos/2019/11/18/chileans-learned-the-right-lessons-after-the-pinochet-era/
Freedom House. n.d.a. Freedom in the World Report. Accessed March 29, 2023. https://freedomhouse.org/sites/default/files/2023-02/All_data_FIW_2013-2023.xlsx
Freedom House. n.d.b. Qatar: Freedom in the World 2019. https://freedomhouse.org/country/qatar/freedom-world/2019
Ghosh, D. 2019. “Facebook’s oversight board is not enough.” *Harvard Business Review*, October 16. https://hbr.org/2019/10/facebooks-oversight-board-is-not-enough
Global Change Data Lab. 2022. “Our World in Data.” Retrieved March 4, 2022. https://ourworldindata.org/grapher/human-rights-scores
Human Rights Watch. 2006. “Sweden violated torture ban in CIA rendition: Diplomatic assurances against torture offer no protection from abuse.” November 9. https://www.hrw.org/news/2006/11/09/sweden-violated-torture-ban-cia-rendition
Human Rights Watch. 2018. “Hiding behind the coalition: Failure to credibly investigate and provide redress for unlawful actions in Yemen.” August 24. https://www.hrw.org/report/2018/08/24/hiding-behind-coalition/failure-credibly-investigate-and-provide-redress-unlawful
Human Rights Watch. 2020a. “Chile: Events of 2019.” *World Report 2020*. https://www.hrw.org/world-report/2020/country-chapters/chile
Human Rights Watch. 2020b. “Columbia: Events of 2019.” *World Report 2020*. https://www.hrw.org/world-report/2020/country-chapters/colombia
Human Rights Watch. 2020c. “France: Events of 2019.” *World Report 2020*. https://www.hrw.org/world-report/2020/country-chapters/france
Human Rights Watch. 2020d. “Peru: Events of 2019.” *World Report 2020*. https://www.hrw.org/world-report/2020/country-chapters/peru
Human Rights Watch. 2021. “Columbia: Events of 2020.” *World Report 2021*. https://www.hrw.org/world-report/2021/country-chapters/colombia
Human Rights Watch. 2023. “Peru: Events of 2022.” *World Report 2023*. https://www.hrw.org/world-report/2023/country-chapters/peru
Huntington, S. P. 1957. *The Soldier and the State: The Theory and Politics of Civil-Military Relations.* Cambridge: Harvard University Press.
ICRC Database. June 8 1977. Protocol Additional to the Geneva Convention of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflict (Protocol II). *International Humanitarian Law Databases.* https://ihl-databases.icrc.org/en/ihl-treaties/apii-1977?activeTab=1949GCs-APs-and-commentaries
Janowitz, M. 1960. *The Professional Soldier: A Social and Political Portrait*. Glencoe: Free Press.
Katzman, K. 2020. “Qatar: Governance, security, and U.S. policy.” *Congressional Service Research Report*. March 11. R44533, Version 62. https://crsreports.congress.gov/product/pdf/R/R44533/62
Kempf, E., and J. M. Shapiro. “Compass Ethics: Governing Through Ethical Principles at WeCorp Industries.” Harvard Business School Case 224-105, July 2024. (Revised December 2024.) http://hbr.org/product/Compass-Ethics--Governing/an/224105-PDF-ENG.
Kutz, C. 2000. *Complicity: Ethics and Law for a Collective Age.* Cambridge: Cambridge University Press.

Lake, D. A. 2001. "Beyond Anarchy: The Importance of Security Institutions." *International Security* 26(1): 129-160.
Levitsky, S., and L. A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes after the Cold War*. Cambridge: Cambridge University Press.
McEachrane, M. 2018. "Universal human rights and the coloniality of race in Sweden." *Human Rights Review* 19: 471-493.
McMahan, J. 2005. "Just Cause for War." *Ethics and International Affairs* 19: 1-21.
McMahan, J. 2011. *Killing in War.* Oxford: Oxford University Press.
Miller, S. 2018. *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction.* SpringerBriefs in Ethics. Cham, Switzerland: Springer.
North, D. C., J. J. Wallis, and B. R. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.
Przeworski, A., S. C. Stokes, and B. Manin, eds. 1999. *Democracy, Accountability, and Representation*. Cambridge: Cambridge University Press.
Redacted A
Redacted B
Repucci, S. 2020. "Freedom in the World 2020: A leaderless struggle for democracy." *Freedom House*. https://freedomhouse.org/report/freedom-world/2020/leaderless-struggle-democracy
Repucci, S., and A. Slipowitz. 2022. "The global expansion of authoritarian rule." *Freedom House*. https://freedomhouse.org/report/freedom-world/2022/global-expansion-authoritarian-rule
Schnakenberg, K., and C. J. Fariss. 2014. "Dynamic patterns of human rights practices." *Political Science Research and Methods* 2 (1): 1–31.
Sen, A. 1999. *Development as Freedom*. New York: Alfred A. Knopf.
Short, J. L. 2013. "Self-regulation in the regulatory void: 'Blue moon' or 'bad moon'?" *The Annals of the American Academy of Political and Social Science* 649: 22-34.
Simmons, B. A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.
Simons, G., and A. Manoilo. 2019. "Sweden's self-perceived global role: Promises and contradictions." *Research in Globalization* 1 (1): 100008.
Skerker, M., T. Lechterman, and B. Strawser. N. d. "The Duty and Limited Permission to Discriminate against Certain Customers." Working paper.
Strawser, B. J. 2023. *The Bounds of Defense: Killing, Moral Responsibility, and War.* Oxford: Oxford University Press.
United Nations (General Assembly). Opened for signature December 16, 1966. International Covenant on Civil and Political Rights. *United Nations Treaty Series*, vol. 999.
United Nations (General Assembly). Adopted on July 17, 1998. Rome Statute of the International Criminal Court. *United Nations Treaty Series,* vol. 2187.
United Nations (General Assembly). New York, 2 April 2013. Arms Trade Treaty. *United Nations Treaty Series*, vol. 3013.
United Nations (General Assembly). 2019. "Visit to France: Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism." A/HRC/40/52/Add.4 https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/133/99/PDF/G1913399.pdf?OpenElement
United Nations (General Assembly). 2020. "Visit to Qatar: Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related." A/HRC/44/57/Add.1 https://documents-dds-ny.un.org/doc/UNDOC/GEN/G20/103/74/PDF/G2010374.pdf?OpenElement

Vivanco, J. M. 2019. “New documents raise fears of a return to ‘false positive’ killings.” *Human Rights Watch*, July 18. https://www.hrw.org/news/2019/07/08/new-documents-raise-fears-return-false-positive-killings

Walzer, M. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

Wettstein, F. 2010. “The duty to protect: Corporate complicity, political responsibility, and human rights advocacy”. *Journal of Business Ethics*, 96 (1): 33–47.

Whetham, D. 2010. “The Just War Tradition: A Pragmatic Compromise.” In *Ethics, Law and Military Operations*, edited by D. Whetham, 65-89. Basingstoke: Palgrave Macmillan.

Winner, L. 2014. “Technologies as forms of life.” In *Ethics and Emerging Technologies*, edited by R. L. Sandler, 48-60. London: Palgrave Macmillan.