ORIGINAL RESEARCH

The technology triad: disruptive AI, regulatory gaps and value change

Jeroen K. G. Hopster[1] · Matthijs M. Maas[2]

Abstract

Disruptive technologies can have far-reaching impacts on society. They may challenge or destabilize cherished ethical values and disrupt legal systems. There is a convergent interest among ethicists and legal scholars in such "second-order disruptions" to norm systems. Thus far, however, ethical and legal approaches to technological norm-disruption have remained largely siloed. In this paper, we propose to integrate the existing 'dyadic' models of disruptive change in the ethical and legal spheres, and shift focus to the relations between and mutual shaping of values, technology, and law. We argue that a 'triadic' values-technology-regulation model—"the technology triad"—is more descriptively accurate, as it allows a better mapping of second-order impacts of technological changes (on values and norms, through changes in legal systems—or on legal systems, through changes in values and norms). Simultaneously, a triadic model serves to highlight a broader portfolio of ethical, technical, or regulatory interventions that can enable effective ethical triage of—and a more resilient response to—such Socially Disruptive Technologies. We illustrate the application of the triadic framework with two cases, one historical (how the adoption of the GDPR channeled and redirected the evolution of the ethical value of 'privacy' when that had been put under pressure by digital markets), and one anticipatory (looking at anticipated disruptions caused by the ongoing wave of generative AI systems).

1 Introduction

Emerging technologies such as artificial intelligence are engines of social change. Such change can manifest itself directly in a range of domains (healthcare, military, governance, industry, etc.) [1]. For instance, technologies can drive shifts in power relations at the societal level [2, 3], as well as internationally [4–6]. Less visibly but no less significant, technological change can also have "soft impacts" [7], by

challenging and changing entrenched norms, values, and beliefs [8, 9]. In virtue of such societally disruptive "second-order effects" [10]—which go far beyond the domain-specific changes of "first-order" market disruptions [11]—emerging technologies such as AI have been described as "socially disruptive" [12] or "transformative" [13].[1]

For instance, while there is still considerable uncertainty over AI technology's future trajectory, AI experts expect continued progress towards increasingly capable systems [14–16]. Further capability developments are likely to make AI's eventual societal impacts considerable, possibly on par with previous radical and irreversible societal transformations such as the industrial revolution [13, 17]. Even when assuming a baseline scenario that (implausibly) assumes no further progress in AI, the mere proliferation of many existing AI techniques to existing actors, and its integration with pre-existing digital infrastructures will suffice to

---

Equal contributions; authorship order has been randomized.

✉ Jeroen K. G. Hopster
  j.k.g.hopster@uu.nl

  Matthijs M. Maas
  matthijs.maas@legalpriorities.org; mmm71@cam.ac.uk

1 Utrecht University and the NWO Gravitation Programme "Ethics of Socially Disruptive Technologies", Utrecht, The Netherlands

2 Legal Priorities Project and the Centre for the Study of Existential Risk (University of Cambridge), Copenhagen, Denmark

---

[1] The concepts of "disruption" and "transformation" are closely related, but do not co-extend. One difference is their respective relation to change: while "technological transformation" foregrounds change, "technological disruption" foregrounds a loss of orientation with respect to some prior state.

Table 1 Key concepts

Disruptive change	Change that instills normative uncertainty and disorientation and calls for reflection on established frameworks of ethics and governance [12]
First-order problems/challenges	The 'easy' problems of AI governance [35], which can be adequately addressed with appeal to pre-existent ethical principles, norms, values, codes, regulations, law
Second-order impacts/disruptions	The 'hard' problems of AI governance [35], which call into question the adequacy of pre-existent ethical principles, norms, values, codes, regulations, law
Technolegal disruption	Technological change that provokes uncertainty over the applicability of existing law [36]
Technomoral change	The two-way dynamic shaping of technology and morality [7]

drive extensive societal impacts [18, pp. 56–82]. Indeed, Dafoe has argued that AI's transformative implications may be grasped by considering it as the next step in a long line of 'information technologies' broadly conceived, spanning back to other such 'technologies' such as speech and culture, writing, the printing press, digital services, communications technologies; or as the next 'intelligence technology', following previous mechanisms such as "price mechanisms in a free market, language, bureaucracy, peer review in science, and evolved institutions like the justice system and law" [19]. Accordingly, we take AI to be a paradigmatic example of an emerging Socially Disruptive Technology [12]—i.e., a technology with the potential to affect important pillars of human life and society, in a way that raises perennial ethical and political questions [20].

The rise of AI has provoked increasing public concern with the technology's potential ethical impacts [21–23], which has translated into growing calls for regulation and ethical guidance [24]. The European Commission has begun to draft an "Artificial Intelligence Act" [25]; Chinese government bodies have articulated new regulatory moves for AI governance, setting out requirements for algorithmic transparency and explainability [26]. There have also been notable steps in governance for AI at the global level [27–29], such as, among others, the establishment of the 'Global Partnership on Artificial Intelligence' (GPAI) [30], or the UNESCO 'Recommendation on the Ethics of Artificial Intelligence' [31], the first such global agreement. Such initiatives reflect the growing view that the sociotechnical impacts of transformative AI should not be left to run their own course without supervision [32], but may require intervention and accountability to safeguard core values such as justice, fairness, and democracy [33]. Yet, scholars, policymakers and the public continue to grapple with questions over how AI is concretely impacting societies, what values it impinges upon, and in what ways these societies can and should best respond.

One challenge to the formulation of adequate responses to the 'first-order problems' posed by AI, is that these can be derailed or suspended by underlying second-order disruptions of this technology, in the foundations and normative categories of both ethics and law (see Table 1 for key

concepts). We understand first-order problems as those that can be adequately addressed in terms of pre-existing norms or prescriptions, such as pre-existing ethical norms or legal codes. For instance, the question of how pre-existing standards of jus in bello can be applied to warfare with autonomous weapons systems is a first-order problem. Second-order problems or disruptions, by contrast, call into question the appropriateness or adequacy of existing ethical and regulatory schemas. For instance, it has been argued that autonomous weapons systems create responsibility gaps that make the very idea of jus in bello inapplicable [34], and it is not obvious how this problem should be resolved. Given its major societal impact, it seems very likely that AI will drive second-order disruptions of various kinds, affecting ethical norms and values as well as systems of regulation. How can we rely on ethical and regulatory frameworks to cope with emerging technologies, while at the same time these frameworks are themselves changed by technology?

In this paper, we propose a conceptual approach that helps to mitigate this challenge, by addressing the disruptive implications of emerging technologies on ethics and regulation in tandem. To date, the fields of Technology Ethics (TechEthics), and Technology Law (TechLaw) have developed sophisticated frameworks that explore the co-evolutionary interaction of technology with existing (moral or legal) systems, in order both to analyze these impacts, and to normatively prescribe appropriate responses. However, these frameworks have remained isolated from one another, and insufficiently acknowledge that norms of TechEthics and regulations of TechLaw co-evolve. We propose to integrate the dyadic models of TechLaw and TechEthics, to shift focus to the triadic relations and mutual shaping of values, technology, and regulation. We claim that a triadic values-technology-regulation model is more descriptively accurate, and serves to highlight a broader portfolio of ethical, technical, or regulatory interventions that can enable effective ethical triage of Socially Disruptive Technologies.

We spell out this claim in the subsequent sections of this paper. In Sect. 2, we further clarify what second-order disruptions amount to and how they challenge TechEthics and TechLaw. In Sect. 3, we present succinct mappings of the dyadic models of TechEthics and TechLaw and subsequently

point out some of their limitations. Specifically, we zoom in on AI technology and explain why second-order disruptions by AI cannot be easily captured by the dyadic models. In Sect. 4, we sketch a triadic model (the "Technology Triad") that aims to synthesize these two frameworks, showing how it helps to grapple with the second-order societal impacts of AI both analytically and prescriptively. In Sect. 5, we evaluate this model, arguing that it can be both more descriptively accurate (as it allows the mapping of second-order impacts on values and norms, through changes in legal systems—or on legal systems, through changes in values and norms), as well as instrumentally useful (and normatively valuable) in responding to these changes, than each of the dyadic models used in isolation. We accordingly provide a step-by-step operationalization of this framework through a series of questions that can be posed of historical, ongoing, or anticipated technology-driven societal disruption, and we illustrate the application of this framework with two cases, one historical (how the adoption of the GDPR channeled and redirected the evolution of the ethical value of 'privacy' when that had been put under pressure by digital markets), and one anticipatory (looking at anticipated disruptions caused by the ongoing wave of generative AI systems). We conclude that approaching disruptive AI through the lens of the "Technology Triad" can lead to more resilient ethical and regulatory responses.

---

² Most famously Moore's law [37], [38]. Moreover, for specific technological subsets (such as defense technologies), technology forecasts have reportedly achieved reasonable accuracy even over several decades [39], though this might be explained by the very long procurement timelines of modern major weapon systems. For a review of the historical efficacy of various attempts at long-range forecasting, see [40].

³ Prediction may be particularly difficult in the domain of AI technology [41]—though that does not necessarily mean that forecasts must always err on the side of excessive optimism [18, pp. 69–71]. For a recent research agenda to forecast advanced AI, see [42]. More generally, attempts to forecast or estimate the development pathways and timelines of advanced AI have been diverse, and have drawn from various lines of evidence or argument, including (listed in order from more abstract to more empirical): (1) philosophical arguments and anthropic reasoning (i.e., from the prima facie likelihood that we would be the ones to find ourselves living in the 'most important century' that contains transformative technologies) [43]; (2) extrapolating general historical trends such as analysis of long-run economic history [44], [45], or the acceleration in the macrohistorical pace of technology developments [46]; (3) estimating specific future development trends by comparing the historical (and likely future) efforts and investments dedicated to creating advanced AI to the amount of resources that eventually proved necessary for major breakthroughs in other scientific fields such as mathematics [47], or by comparing the (comparatively) limited past investments in AI to the likely growing future resources dedicated to this field [48]; (4) estimates based on meta-induction from the (good or bad) track record of past technological predictions, especially those made by futurists [49], [50]; (5)

2 Background: technological change and first- and second-order disruptions to ethics or law

The pace of policy responses to disruptive technological changes tends to be relatively slow, which may be due to various reasons. One factor is the uncertainty over the future course of the technology's sociotechnical trajectory. With some notable exceptions,² it has usually proven difficult to accurately predict a technology's future societal uptake in advance.³ It is often even more difficult to anticipate a technology's subsequent uptake and use in society, let alone the resulting societal impacts [62, 63]. As such, there may often be legitimate disagreement about the costs and benefits of adopting either a permissive or a precautionary approach towards regulation [64].

However, there is another barrier as well, which pertains to the ways in which technological disruption can stress the normative credentials of existing ethical heuristics and the functioning or legitimacy of available regulatory response strategies. Socially Disruptive Technologies can have "deep impacts" in ethics and beyond [12]: they transform basic ethical concepts, norms, and public values, which are instrumental to the ethical assessment and guidance of emerging technologies in turn. For example, it has been argued that two core human values—"truth" and "trust"—are being disrupted by emerging information technologies, yielding new norms of veracity and trustworthiness [65]. As a result, ethicists face a challenge: should they rely on prior norms and conceptions of truth and trust, or should they rethink these, in responding to the challenges of disruptive technologies?

Emphasis on second-order impacts of technologies is given in the field of ethics that studies emerging technologies (henceforth TechEthics); this literature often frames such shifts in terms of 'technomoral change' [7, 66–68]. The core premise of the technomoral change lens in TechEthics is that ethics and technology evolve in mutual interaction and shape each other: technological artifacts and applications are frequently designed to reflect and realize social and moral values,⁴ but technologies may end up reshaping and

---

Footnote 3 (continued)

surveys of specialists (AI experts) expectations of progress [14], [15], [51], [52]; (6) estimates based on generalist forecaster predictions [53–55]; (7) first-principles analysis, such as a comparison of projected trends in falling costs of training AI systems, against the minimum amount of computation needed to recreate human biological cognition [56–58], among many others. For a survey of methodologies, see [59], and for accessible overview and discussions of various approaches to forecasting the development (timelines) of advanced AI, see [16], [60], [61].

⁴ This aspiration is explicitly hailed by "value sensitive design" [69], [70] and "ethics by design" [71] approaches in fields of engineering and responsible innovation. But the flipside of value-embeddings

disrupting our norms and values in turn. While technomoral change is not the only approach adopted in the field of Tech-Ethics [73], it is certainly a prominent approach, especially for anticipating the implications of emerging technologies.

Analogous discussions on the second-order disruption of established norms occur in the field of law and regulation. Here, the challenge to simply applying existing law to address new (but essentially familiar) first-order problems created by emerging technologies is that often, the features or uses of these new technologies do not lend themselves to easy categorization, provoking legal uncertainty [64, 74]. For instance, cryptocurrencies blur the lines between different types of more traditional assets; different regulators have classified them as a currency, a security, or a commodity [64]. Such classificatory challenges are sometimes framed as being driven by the alleged 'novelty' of a technology: the 'exceptionalist' argument here is that some new artifacts, or some of their uses, are so different from past technologies, that the existing laws cannot sensibly or reasonably interpret and decide upon the new situation.[5]

Given the potential inflexibility of law, it has been frequently argued that the speed and complexity of emerging technologies creates a 'pacing problem' for regulatory and governance responses [83, 84], though some have critiqued this concept [85]. However, the more recent approach of 'TechLaw' [64] does not grant the premise that the law can never keep up with technology. Instead, TechLaw focuses on "how law and technology foster, restrict, and otherwise shape each other's evolution" [64, N. 1], [86]. Confronted with technological changes that affect legal rules, the legal system may respond in three ways: (1) by trying to deal with the new technology under existing rules (often through analogy to previously regulated technologies or their afforded behaviors), as occurs, for instance, when autonomous weapon systems are analogized to other weapons and regulated under existing weapon law; (2) by extending or modifying existing rules to fit the new technology, as occurs, for instance, when U.S. copyright law which restricts unauthorized copying "by any method now known or later developed" is extended to new technologies; (3) by creating new rules [64, 87], as exemplified by the new "AI Act" currently developed by the European Parliament. There is no one-size-fits all answer as to which response fits best; instead, the primary aim of the TechLaw approach is to identify how familiar forms of legal uncertainty appear in new sociolegal contexts [64, 88, 89].

In sum, the existing TechLaw and TechEthics approaches both already recognize and foreground the evolutionary nature of their respective domains: both law and ethics are understood as not static but rather evolving systems, which take their shape in interplay with a variety of (first-order and second-order) pressures, technology prominent among them. However, while the evolutionary nature of both law and morality has been recognized in recent scholarship, another shared feature of TechEthics and TechLaw has remained obscured: that (technology) ethics and (technology) law are co-evolutionary systems with mutually dependent trajectories. That is, while both scholars of morality and law have been paying increasing attention to the interrelations of their fields with technology, they remain largely oblivious to the entangled dynamics of their fields with one another.[6] This

---

Footnote 4 (continued)

has also been acknowledged by philosophers of technology, for in instance in the classic work of Winner [72].

[5] Such challenges of legal uncertainty can frequently be seen in the recurring legal debates over the latest generation of 'emerging technologies', defined as technologies that display a range of attributes, such as (i) radical novelty, (ii) relatively fast growth, (iii) coherence, (iv) prominent impact, and (v) uncertainty and ambiguity [75]. Emerging technologies are considered particularly challenging for regulation to deal with: they may create uncertainty over whether or how existing law applies [76]; they might fall into a gap between pre-existing institutional mandates [77]; and any debates to resolve these uncertainties may be held up by new political challenges or tensions. Within domestic law, the deep ethical impacts referred to above may inhibit easy resolution of the legal uncertainty; within international law, scientific information inequalities and outlying political disagreements may likewise often lead to gridlock, inhibiting the updating or creation of needed regimes [78]. This may even lead to the gradual erosion of existing norms, and the obsolescence of existing treaty regimes [79], creating so-called 'jurisprudential space junk' [80]. However, see Eichensehr [81] for an argument that holds that, while the arrival of new technologies on the international stage frequently prompts debate over whether they are covered by existing international law (the 'international law step zero'), there are several factors that contribute to these uncertainties often being resolved in favor of existing norms applying to the new technologies (rather than prompting an outright legal gap and need for new law). See also Israel's [82] discussion of how soft law development can contribute to the evolution of governance, even in conditions of apparent 'treaty stasis' when hard law regimes cannot be updated.

[6] This holds, in particular, for the dominant analytical models used to anticipate how TechEthics and TechLaw evolve and how to intervene in them. We acknowledge, however, that there are general approaches to the guidance and regulation of technology that do recognize an overlap between ethics and law. So-called ELSI (US) or ELSA (Europe) studies—studies of Ethical, Legal and Social Aspects of scientific and technological developments—have received substantial funding since the 1990s [90]. ELSA was succeeded, in Europe, by the Responsible Research and Innovation (RRI) programme, which has been actively promoted by the European Commission [91], [92], alongside other Responsible Innovation (RI) initiatives. With its focus on the social responsibility of innovators and other stakeholders, as well as the governance of technological innovation, RRI, too, explicitly approaches technology at the intersection of ethics and law. The same, we should add, holds for many of the recently developed frameworks for the ethical guidance and legal regulation of AI [93], which typically outline combined ethical and legal instruments and procedures, and are often embedded in an overarching RRI framework.

may not be a pressing problem when it comes to analyzing and responding to small-scale technological disruptions, but we will argue that second-order disruptions by AI require a more integrative approach.

3 Two dyadic models: TechEthics and TechLaw

Before arguing for the benefits of an integrative triadic model (Sect. 4), let us first outline the background of the existing dyadic models used in TechEthics and TechLaw, starting with the approach of technomoral change. The core premise of this approach is that ethics and technology mutually shape each other. The emergence of contraceptive technologies provides one of several historical case-studies illustrating this mutual shaping: the invention of the female birth-control pill was driven by social activists pursuing a variety of social and moral goals [94], but at the same time, by severing the link between sex and pregnancy, the birth-control pill facilitated unanticipated shifts in the sexual morals of many societies, and fueled emancipation movements far beyond the expectations of the initial reformers. Various other historical examples have been discussed in the literature on technomoral change, such as the influence of ploughing technology on gender norms; the dynamics between new weapon technology and the demise of dueling as an exclusively aristocratic practice; or the role of veterinary medicine and meat replacements in changing attitudes towards the treatment of farm animals [95].

While these historical cases provide a proof of concept, the technomoral change framework is mostly used as an anticipatory framework, which serves to sketch scenarios of possible pathways of future value change. This 'technomoral scenario approach' has recently been extended with the approach of 'axiological futurism', which proposes a systematic exploration of future axiological trajectories [8, 96]. These anticipatory frameworks are part of a broader array of Ethical Foresight Approaches [97], which ethicists invoke to assess emerging technologies. Frequently, these approaches involve a combined effort of not only anticipating the future dynamics of change, but also assessing change in prescriptive terms and intervening to achieve desired outcomes.

A recent criticism of technomoral change is that the approach faces an explanatory gap [68]: it does not provide a clear explanation as to why some technomoral changes have a decidedly disruptive character. Sometimes technology and morality shape each other gradually; at other times, changes occur rapidly, unleashing powerful disagreement and confusion. Nickel et al. [68] argue that this explanatory gap can be filled by providing a more comprehensive account of what moral inquiry and moral change amounts to, which should emphasize the role of individual and collective moral

uncertainty and confusion about the interpretation, priority and correct application of public values.

Adding to this, we submit that a more comprehensive account of technomoral change should integrate with TechLaw work on (techno)legal disruption [36, 74], giving recognition to the legal and regulatory uncertainty that accompanies technological disruption. When legal and regulatory systems are disrupted, it is not obvious which existing legal and regulatory frameworks, if any, apply to a technology; or (if they are still held to apply), it is not obvious how they will apply. Such legal uncertainty, in turn, loosens a potential constraint on technomoral change: in the absence of regulatory standards, the dynamics of future technomoral change are more difficult to anticipate. Consider the example of generative AI, which we will discuss in further detail in Sect. 4: in the absence of standards for AI regulation, the question of how AI will affect societal norms and values is much more open-ended, than when such standards are present. Legal and regulatory gaps and a loss of institutional bearings loosen the bounds of collective moral inquiry, whereas the presence of a regulatory framework imposes a (soft) constraint on it.

A further criticism is that scholarship in TechEthics barely touches on questions of radical moral change at a societal level. The thematic focus of current case-studies in this literature is somewhat narrow and primarily geared to the biomedical sphere, as suggested by the frequently used example of the birth-control pill [98, 99]. Furthermore, research on the co-shaping of technology and society more generally is often geared to interactions between humans and apparently mundane technological artifacts. In both of these respects, we submit, extant scholarship is not perfectly equipped to analyze the deeper impacts of Socially Disruptive Technologies. Like AI, these are often not artifacts but sociotechnical systems; furthermore, the object of study is typically more radical forms of societal change.

Next, let us consider the field of law, regulation and technology, where there has similarly been a sustained focus on the mutual shaping of emerging technologies and particular regulatory systems [100]. Such work has frequently focused on the legal impacts of one or another specific (anticipated) new technology—from new reproductive technology to nanotechnology, and from the internet to AI applications—on existing law or doctrines [87]. Often, these debates have turned on the perceived 'novelty' of the technology under question, or of its assumed 'essential characteristics'.[7]

---

[7] This led some to assert that no new technologies would ever be so problematic or disruptive to existing law—as with the dismissal by Judge Easterbrook that new laws distinctly tailored for a new technology (specifically cyberspace) would be as superfluous as attempting to specify a separate 'law of the horse' [101], since existing bodies of law would be more than flexible enough to be extended to situations involving a new technology. Other scholars responded by asserting that there were in fact distinct legal challenges posed by the internet

Accordingly, such legal work drew on an exceptionalist approach, asking whether or when a particular new technology possessed sufficiently novel or remarkable 'essential features' that it cannot be adequately covered by the existing legal doctrine.

Recent legal scholarship has taken issue with this exceptionalist framing of technolegal disruption, arguing that disruptive technologies foreground familiar forms of legal uncertainty in new sociolegal contexts [64, 88]. Relatedly, some scholars have called to depart from technology-centric or application-centric approaches to regulation [105, 106], and to instead focus on general types of change in the regulatory ecosystem. What matters in this view are not the assumed artefactual characteristics of a technology, but rather the societal 'salience' [104] or sociotechnical changes [88, 106] resulting from its use. Such work has sought to take a more systematic approach to developing general frameworks for understanding the cross-sector 'legal disruption' of technology [36].

In sum, both fields—TechEthics and TechLaw—have provided important insights into processes of technomoral change and technolegal disruption, respectively. As dyadic models, they each improve upon older approaches by allowing for an analysis of the mutual shaping of both phenomena under examination (technology and ethics; and technology and law, respectively). But both models, in isolation, are not ideally suited to anticipate and assess the implications of Socially Disruptive Technologies.

3.1 AI and second-order disruptions through dyadic lenses

Let us outline the strengths and shortcomings of the current dyadic models, with a focus on the case of AI. In recent years, AI has been adopted in diverse practices, from targeted advertising, insurance pricing, and fraud detection, to hiring decisions, predictive policing, and administrative decision-making. Notwithstanding the various benefits of the technology, it has also been associated with concerns about discrimination, privacy infringement, and the spread of misinformation [107], among many others. AI systems have been described as potentially causing harm at many levels—individual harm, collective harm, or societal harm [108]. With no aspiration to being exhaustive, we can represent some of the concerns that AI raises in terms of the dyadic models of both TechEthics and TechLaw, in terms of the mappings of Figs. 1 and 2.

The starting point of the dyadic TechEthics model is to analyze ethical problems and disruptions emerging to which emerging technologies give rise, and to subsequently identify which ethical response might—or should—be prescribed. As Fig. 1 illustrates, this dyadic approach allows us to examine important pathways by which technologies can lead to new first-order challenges and changes. For instance, TechEthics scholarship can identify and analyze numerous cases where new (AI) technology creates first-order ethical problems, because it violates established and cherished public values, such as privacy, non-discrimination, democracy [109, 110], human dignity [111], or environmental sustainability [112].

Where first-order challenges are concerned, these public values remain stable: while AI challenges compliance with extant ethical norms and practices (of non-discrimination, privacy, etc.), first-order challenges do not include a more thoroughgoing contestation of norms of privacy or the value of non-discrimination. Yet, the dyadic TechEthics approach can also reckon with such second-order ethical disruptions: it can analytically identify second-order changes in value systems in interaction with technological changes. On the basis of these analyses, the model allows the prescription of appropriate responses (of the ethical value system) to both first-order problems and second-order disruptions. Hence, the model is dyadic: it also includes the reverse, prescriptive question of how (ethics or law) can and should shape technology.

While useful for many contexts, a shortcoming of the dyadic TechEthics model is that it makes no explicit reference to regulation. As a result, in terms of analyzing the dynamics of technomoral change, it can easily miss out on relevant mutual interactions, as well as on key underexplored pathways, such as indirect effects of technology on ethics that are mediated through intermediate effects in the domain of law. In terms of recommendation, it risks foregrounding some prescriptive responses over others. In particular, there is an inclination to focus on interventions that can be made by altering the technology—either to the artifact, or its design process, as seen in 'value-by-design' approaches [69]. Conversely, regulatory responses are de-emphasized.

Now consider the dyadic model employed in TechLaw scholarship (Fig. 2). This model allows researchers to analytically identify and characterize a range of domains where AI systems put pressure on the existing regulatory equilibrium. This can be because the use of AI can create first-order problems that raise the question of whether or how existing laws apply, as seen in situations ranging from the lack of clarity over the status of self-driving cars under international road traffic conventions [113]; whether 'robot lawyers' should be classified as 'goods' or 'services' in international trade law [114]; whether autonomous weapons systems violate the norms of International Humanitarian Law [115]; or whether

---

Footnote 7 (continued)

[102], leading to the development of distinct technology-specific legal theories such as 'cyberlaw' and 'robolaw' [103, 104].
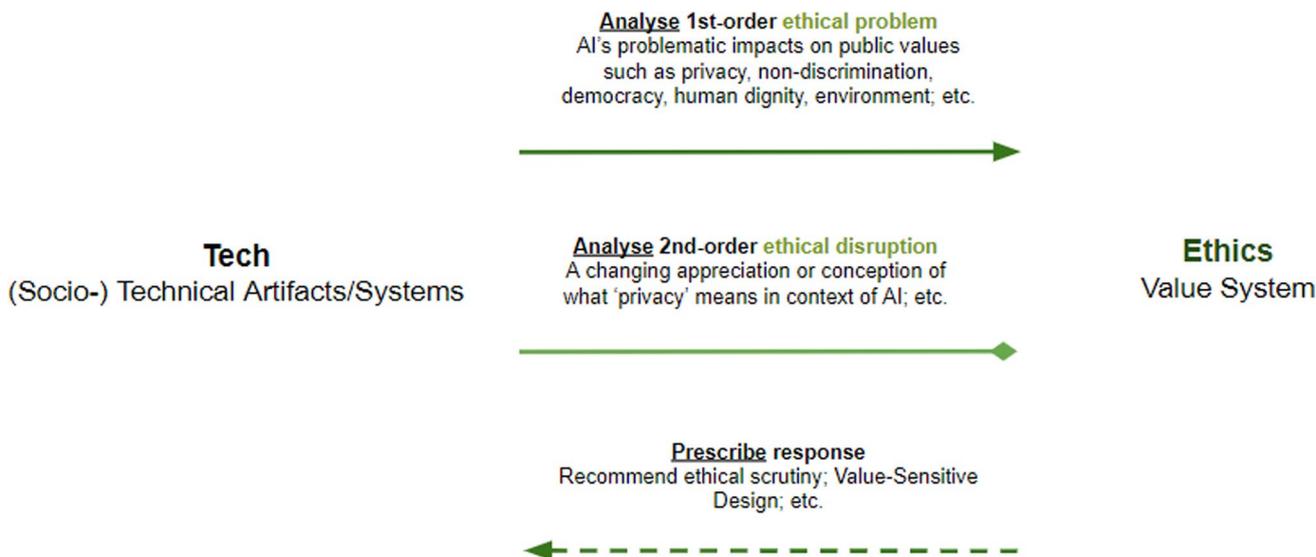
Fig. 1 The dyadic approach to Tech Ethics
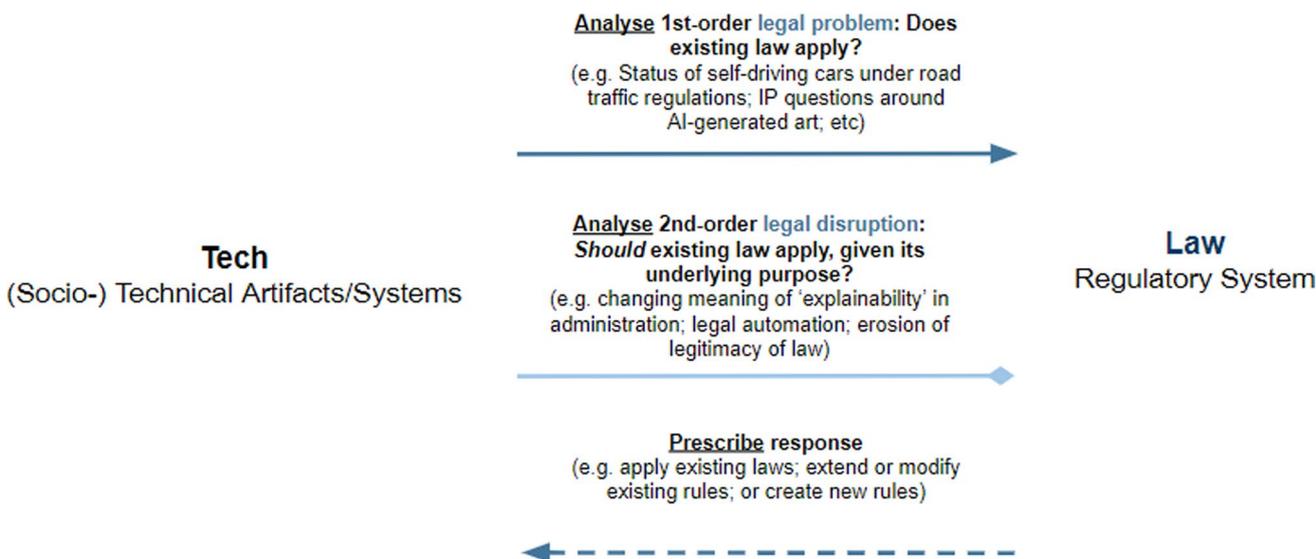
Fig. 2 The dyadic approach to TechLaw

existing criminal law doctrines such as 'mens rea' can neatly apply to 'crimes' carried out or commissioned by AI systems (such as trading agents convergently discovering fraudulent trading strategies) [116, 117].

In principle, TechLaw scholarship can also explore cases of second-order legal disruption, where new developments create uncertainty over whether existing laws should apply, or whether they should rather be reconfigured in light of the new situation [64]. But note that these normative considerations cannot be settled based on intra-legal considerations; instead, they require ethical

reflection on the proper scope of legal and regulatory intervention, and on the benefits and risks of taking a pre-active regulatory stance in the face of Socially Disruptive Technologies. This calls for drawing on ethical principles and values—a component not foregrounded in the dyadic TechLaw model.

In sum, in both dyadic frameworks, we can study and analyze situations where second-order impacts of AI systems cannot be resolved with a straightforward appeal to existing frameworks, but raise uncertainty about—or call

into question—these frameworks themselves, or important aspects of them.[8] Some recurrent issues, both in the fields of TechEthics and TechLaw, pertain to the question of how to keep humans 'in the loop' with the advent of AI [119], to think of new hybrid forms of human–machine agency and responsibility, and to design 'humane' AI technology. Such issues require foundational reflection: rather than a problem that can be solved in the context of the current ethical/legal ecosystem, the ecosystem itself is challenged and may possibly need to be amended or reformed in order to cope with the challenge. In the face of this challenge, the dyadic models of TechEthics and TechLaw encounter limitations. When norms and values themselves are transformed, or when regulatory systems are disrupted by technology at a fundamental, conceptual, or processual level, then an adequate analytical framework requires a more holistic overview of the resultant changes in the combined ethical-regulatory ecosystem, both in order to grasp the relevant dynamics and to respond to them.

4 The triadic model

While the dyadic models of TechEthics and TechLaw have been developed separately and have thus far largely worked in parallel, they share at least part of their analytical domain: they are both concerned with technology (whether particular artifacts or sociotechnical systems),[9] and aim to explore the dyadic, co-evolutionary relation of technology with a particular social system or conceptual order. One focuses on the dyadic relations of technology to ethics (here broadly understood to include social value systems and norms); the other on the dyadic relations to law (i.e., a particular regulatory system).

This means that we can map all three systems onto a simplified triadic map, which allows us to map the overlap and differences between these dyadic paradigms (see Fig. 3). In particular, we can map which types of analytical topics and paths each of them highlights, and what kinds of prescriptive

---

[8] For instance, widespread use AI of may raise questions about the legitimacy of democratic institutions in the face of potential manipulation by AI, about the applicability of existing ethics codes given the new pressure that AI puts on the value of explainability, or about the applicability of existing law to AI. These second-order disruptions are related to what Minkkinen and Mäntymäki refer to as the 'hard' problem of AI governance, which "concerns AI as a general-purpose technology that transforms societies, communities, and potentially even human beings," and which rather than a matter to be resolved "is a sensemaking process regarding sociotechnical change." [35], see also [118].

[9] For definitions of 'technology' and 'sociotechnical systems', see also [18, 76, 120].

evaluations or recommendations either field reasons towards or foregrounds.

4.1 Analytical and prescriptive pathways on the triadic model

We suggest that integrating both models into a single triadic model can offer many more pathways to both analyze societal disruption (see Fig. 4), as well as a wider palette of potential responses available to both fields (see Fig. 5).

The triadic model allows an analytically richer exploration of indirect technological impacts: it can explore the impact of AI on values, as a result of its disruption in legal systems; or on legal systems, as a result of technology-driven changes in underlying public values. As such, the model illustrates the benefits, both descriptive and prescriptive, of the triadic approach over the isolated dyadic approaches.

In terms of descriptive analysis, while, as discussed in Sect. 3, both TechEthics and TechLaw can to some extent explore first-order and second-order challenges, they usually emphasize first-order challenges, and face limitations when exploring second-order challenges. Normative considerations are certainly part of extant TechLaw approaches, but these considerations are not foregrounded if TechLaw is framed in dyadic terms. A shift of focus specifically benefits analysis of second-order disruptions which are mediated indirectly—i.e., which results not directly from technological change, but indirectly from technological change mediated in the other domain. For instance, there might be use cases of AI (sociotechnical developments) that predominantly or most visibly affect regulatory systems, yet which raise important indirect effects for ethics. This could include first-order challenges for ethics (e.g., the accountability implications of increasing automation of legal decision-making), as well as second-order changes in ethics (e.g., how the increasing automation of legal systems might result in a shift in the extent to which society values transparency relative to efficiency or speed in governmental decision-making [121]). The triadic model enables easy identification of such second-order effects.

Simultaneously, the triadic model offers more actionable prescriptive analysis, addressing some of the shortfalls in the dyadic models. In particular, it improves upon dyadic approaches by (a) allowing for a more appropriate analysis of prescriptive priorities—i.e., triage among the full set or spectrum of societal disruption (both to ethics and law) driven by AI technology (see also Sect. 5). Moreover, the triadic model (b) supports a normatively richer analysis of potential prescriptive responses, as it prompts TechLaw and TechEthics to have a fuller appreciation of the relevance of carrying out responses through one another's tools.

4.2 Operationalizing the triadic model

Having sketched the triadic model in the abstract, let us now indicate how it can be operationalized. We do so by outlining three steps. The first step is identificatory: it consists of identifying a relevant case (historical, ongoing, or anticipated) of technology-driven second-order disruption to ethical and/or legal systems. Questions that can help to guide such identification are[10]:

a. Which (past, ongoing, or anticipated) technologies meet the criteria for 'emerging technologies' [75], and/or 'socially disruptive technologies' [12, 122], such that we should expect not just first-order problems but also second-order impacts on ethics and law?
b. Which (past, ongoing, or anticipated) technological disruptions are studied in both TechEthics and TechLaw scholarship, but primarily with a focus on domain-specific first-order disruptions? Where does either lens focus on the (problems created by) new artifacts when below the surface there are larger sociotechnical systems?
c. Which (past, ongoing, or anticipated) technological disruptions are currently identified and studied as second-order impacts in either TechEthics or TechLaw scholarship, but remain understudied and underappreciated in the other?
d. Which (past, ongoing, or anticipated) technological disruptions have received attention from both TechEthics or TechLaw, but generally receive very different treatment, analyses or evaluation?
e. Which (past, ongoing, or anticipated) technological disruptions have received attention from both TechEthics or TechLaw, but both recommend different responses?

The second step consists of reviewing and comparing existing dyadic accounts to analyze second-order disruptions. Adopting the dyadic TechEthics lens: how does TechEthics analyze the technomoral change? Is the ethical shift one of (de)valuation, conceptual reconstitution, or gradual shift in ethical values? In addition, what responses does TechEthics accordingly prescribe? Or adopting the dyadic TechLaw lens: how does TechLaw analyze the technolegal disruption? Does the new artifact or enabled behavior (a) create clear gaps obviously uncovered within existing law; (b) lead to incorrectly over-inclusive or incorrectly under-inclusive application of existing laws; (c) lead to the obsolescence of laws (e.g., because they are no longer needed, adequate, or enforceable); or (d) shift the relative balance of problems?[11] In addition, what responses does TechLaw accordingly prescribe? E.g., when or where does (/should) the legal system respond to the new technology (a) by dealing with it under existing rules (e.g., through analogy); (b) by extending or modifying existing rules to fit the new technology; or (c) by creating new rules?

The third step is to integrate both dyadic accounts into a triadic model. In terms of analysis (3a), this may allow for the identification of legal disruptions that follow (indirectly) from technomoral change, e.g., (i) because the shift in the view or conceptualization of key values, indirectly affects the necessity, legitimacy, or underlying purpose of key existing technology laws, making their (re)application problematic and/or changing their intended purpose; or (ii) because the commonly prescribed ethical responses may create new conflicts or contradictions under existing legal systems. Conversely, technolegal disruptions may give rise (indirectly) to ethical changes, e.g., (i) because the regulatory response to patch legal provisions for the technology itself becomes considered ethically problematic or contested; or (ii) because the regulatory response affects, redirects or channels the public process of technomoral change into different directions.

In terms of triadic prescription (3b), the point of the third step is to identify new priorities, strategies, or considerations for societal (ethical and/or legal) responses to an emerging technology. Three types of use of the triadic model can be distinguished here:

(i) Triaging prescriptive priorities. Taking the broader view of the technology's societal disruptions to both ethics and law, which of these are the most urgent, critical, or fundamentally disruptive? Are these the direct second-order disruptions in either one domain (ethics or law), or are these the indirect second-order disruptions? How should this shift the priorities or research agenda of either TechEthics or TechLaw scholarship?
(ii) Tailoring prescriptive responses within a lens. To TechEthics, what does this triadic perspective highlight about the multiple realizability of ethical responses to the technological disruption? To TechLaw, how can the triadic perspective help make regulation more tailored to the actual societal disruption?
(iii) Tailoring prescriptive responses between lenses: where could either field draw on tools from the others' toolset in addressing the societal challenges it faces?

---

[10] The questions outlined under each of the three steps constitute an indicative, but non-exhaustive list of relevant questions within the triadic approach. Their relevance depends on the case-study at issue.

[11] See again [18, p. 196].
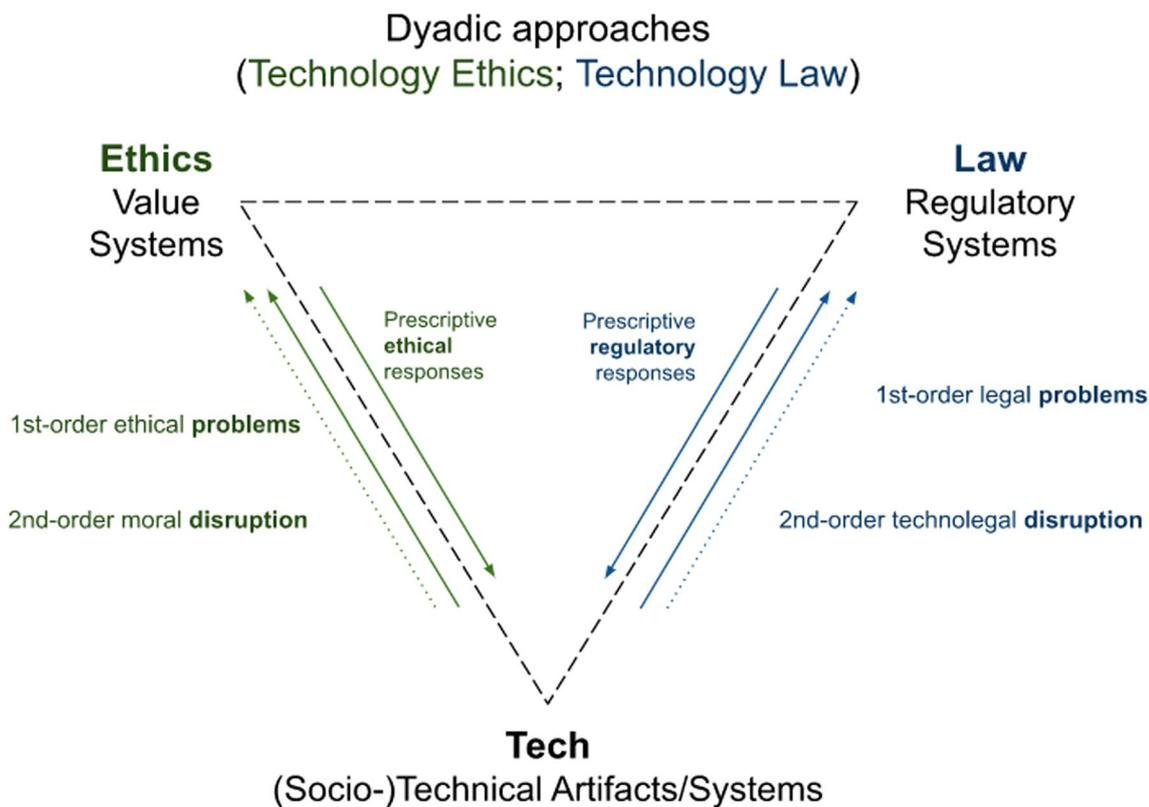
Fig. 3 Siloed dyadic approaches, mapped on the triadic model

Fig. 4 Analytical pathways on a triadic model
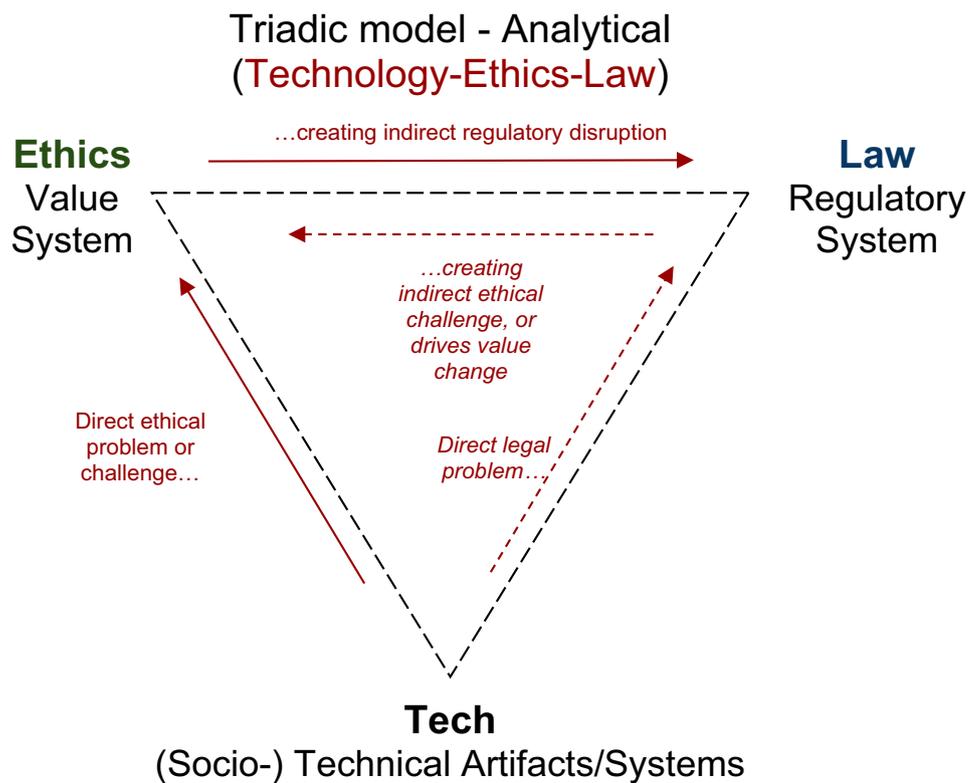
Fig. 5 Prescriptive pathways on a triadic model
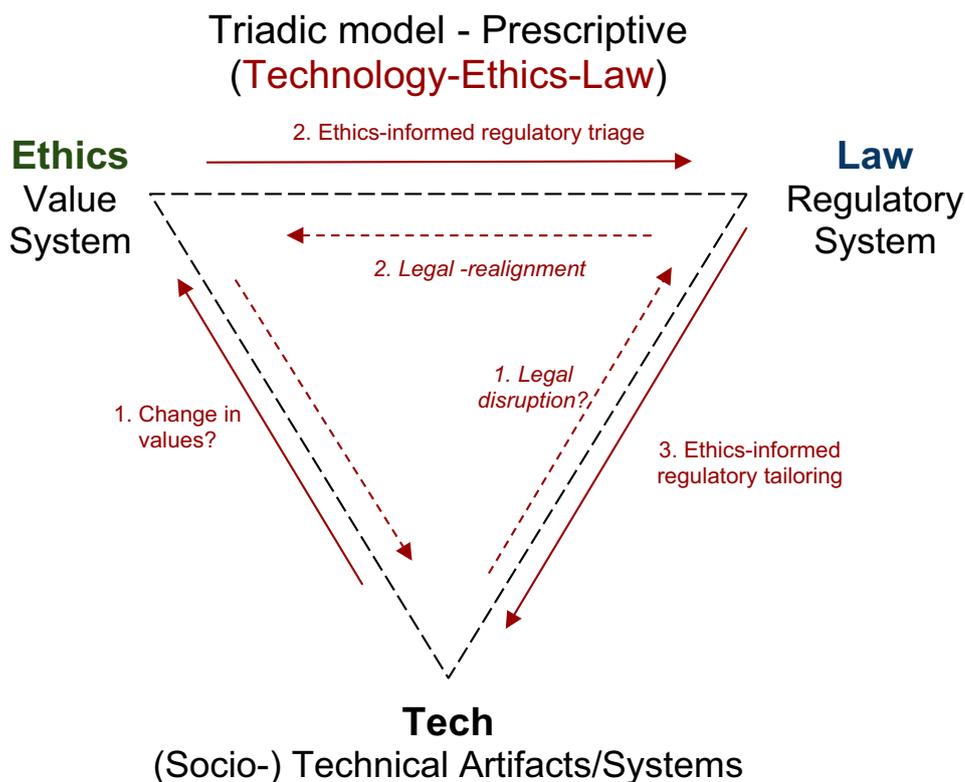
4.3 Illustrating the triadic model: two case-studies

We conclude this section by sketching two case-studies on the basis of this three-step approach.[12] The first case-study is historical: it considers the intersection of digital technology, the societal value of privacy, and (data privacy) regulation within the last two decades. The main upshot of the triadic model, here, is to impose structure and enhance analytic clarity, by following general steps and questions to describe the dynamics of a second-order disruption. Nuanced reassessment of historical cases can often be highly revealing, both in shaping our views on the genealogy of our current (technology-focused) values or laws—our technomoral and technolegal legacy—as well as in providing potentially transferable lessons for how to anticipate novel instances of technomoral change.

The second case-study considers the growing use and dissemination of increasingly general-purpose and 'generative' AI systems. Here, the triadic model can help both to anticipate the relevant dynamics of this still-unfolding process of techno-moral-legal change, and to make recommendations for intervening in them. In particular, the model foregrounds that the emergence of generative AI does not only require scrutiny of the soundness and applicability of existing regulations, but also calls for ethical reflection on the value of human creativity, authenticity, and inventiveness, which should be reflected in the normative assessment of these AI technologies.

4.4 Case-study 1: digital technology as threat to privacy

Step 1: Identify a case of second-order ethical and/or legal disruption.
While the last decades have seen extensive work exploring the ways in which digital technologies have challenged or endangered privacy, their impacts have not just been in creating first-order ethical (or legal) threats to privacy. Rather, a plausible case can be made that over the last 2 decades, privacy norms have also significantly shifted under the pressure of digital technologies, such that some conceptions of privacy—such as an understanding of privacy in terms of "secrecy"—have lost some of their appeal and prominence. This can be understood as a form of second-order ethical disruption.
Step 2: Review and compare existing dyadic accounts to analyze direct second-order disruptions.

---

[12] We think of these as indicative and incomplete treatments of complex cases; later analysis could and should extend this evaluation at far greater length and detail. Our aim, in the limited scope of the present exposition, is to provide an indication of the ways in which a triadic approach to technology, ethics, and law, can be analytically informative as well as prescriptively useful.

A dyadic model of TechEthics might serve to anticipate this technology-precipitated value dynamic. A technomoral change lens might, for instance, portray the dynamic as a runaway effect, resulting in a society which steadily and effectively surpasses privacy. Conversely, a TechLaw lens might focus on how emerging practices of consumer data collection and tracking not only create a need for new regulations to limit such practices of online privacy infringement (first-order legal disruption), but also lead to deeper reassessments by technology lawyers of the legitimate and/or appropriate pathways (such as 'code as law') through which regulators can serve and protect such values [102].

Step 3a: Integrate into a triadic model: analysis and anticipation

In terms of analysis, a triadic lens can help to understand how, at least in a European context, this societal dynamic has arguably been stalled and transformed by regulatory intervention. Specifically, in 2016, the EU adopted the General Data Protection Regulation (GDPR). Its original aim was to replace the older 1995 Data Protection Directive, and to establish new guidelines for a much more connected world (technological change creating a direct legal problem). In response, the GDPR sought to adapt data protection regulations in ways that preserve 'privacy' as a core value (legal response to legal problem). However, rather than aiming to hold on or preserve or restore previous notions of "privacy as secrecy", what the GDPR has functionally safeguarded is often more closely related to a conception of privacy in terms of the "appropriate flow" of information [123] (legal response driving indirect value change). Hence, the impact of technology regulation is closely linked to our retroactive understanding of these recent dynamics of value change; while digital technology is plausibly seen as a key driver of these changing norms and conceptions of privacy, the relevant dynamics are better appreciated by foregrounding the tacit third element of (technology) regulation in channeling and directing the process of technomoral change.

4.5 Case–study 2: the rise of generative AI

Step 1: Identify a case of second-order ethical and/or legal disruption.

The recent and ongoing rise of multimodal and increasingly general 'foundation models' [124–126], large language models (LLMs, e.g., GPT-4, Claude, Bard), generative AI (e.g., DALL-E 2, Stable Diffusion, Midjourney), and other large generative AI systems

[127],[13] has received tremendous attention, and it is widely expected that this is a development that brings tremendous new challenges to both ethics and regulation [128]. To date, there has been significant attention to many first-order problems of these generative models. This includes work on ethical first-order problems, such as these models' potential to produce biased or hate speech, leak sensitive information, produce generally poor quality information, or aid the production of fake news or hate speech [129, 130]. Others have expressed concern over the growing capabilities and risks as these AI systems are scaled up further, and have called for temporary pause in such experiments [131]. Simultaneously, in TechLaw there has been emphasis on legal first-order problems, such as Italy's 2023 ban on ChatGPT, in the wake of privacy concerns, and the intent to evaluate its compliance with the GDPR [132]; copyright lawsuits by artists and coders over the use of open-source materials in training LLMs [133], or Chinese regulations on generative AI to ban 'subversive content' [134], among many others. However, generative AI systems also have a clear potential to yield second-order disruptions, which call into question the applicability of existing concepts or norms.

Step 2: Review and compare existing dyadic accounts to analyze direct second-order disruptions.

Generative AI systems put pressure on existing ethical concepts and intuitions. For instance, generative AI creates a new credit-blame asymmetry when assigning responsibility for language model outputs, in that human users should still be blamed for utilizing bad or low-quality outputs of those systems, yet should not get (as much) credit for utilizing particularly good outputs [135]. Another concern is that the dissemination of generative art models makes it increasingly unclear how to understand and value the notions of creativity and authenticity, as these allow for the reproduction of individual artistic styles to create new artistic products at scale, or the generation of novel texts that can mimic existing writing styles and contents. Moreover, there are cases of direct legal disruption. For instance, since their proliferation in late 2022, the latest generation of generative AI chatbots has rapidly put pressure on the risk-based approach of the proposed EU AI Act [136]; because these general-purpose AI systems have a wide range of possible use cases, they rapidly made it very difficult for the provider to envisage its downstream risk [137]. This highlighted the shortfalls of regulating AI at the application layer rather than throughout the product cycle [138]. In response, some have suggested applying strict liability [139].

Step 3a: Integrate into a triadic model: analysis and anticipation.

---

[13] Within this analysis, we will broadly focus on the term 'generative AI', as this is a term that has received a particular traction in recent TechEthics and TechLaw scholarship. See also adjacent terms ('Large Generative AI Models') in [127].

Where will direct technomoral change create indirect legal disruption? Ethically thick concepts such as 'creativity' 'authenticity', and 'inventiveness' have long served as cornerstone concepts in legal discourse on intellectual property and patent law. Many aim to reapply these concepts to generative AI art. Yet one challenge is that the sheer proliferation of AI systems, resulting in widely available artistic capabilities, may begin to draw into question old ways of valuing creativity; this may re-open debates over whether, or how, IP law should be applied to protect or safeguard particularly human creativity.

A related challenge is that, as LLMs begin to change the nature of many workplace tasks, the aforementioned growing credit-blame asymmetry will begin to express itself in an 'achievement gap', whereby many human jobs will involve supervising, prompting or maintaining LLMs to produce the outputs that skilled humans previously received credit for; but where it becomes increasingly hard for human professionals to claim credit for these tasks [135]; this may lead to a reappreciation of the nature and value of meaningful work, which might be taken as a need for regulatory updating in domains such as labor and employment law.

Where will direct technolegal disruption create indirect ethical change? Some regulatory initiatives may focus on ensuring the 'democratization' and access to new technologies. Yet the ease with which generative AI can be disseminated, yet also misused, appears set to create new ethical debate and renegotiation over what it means to have 'democratised' (AI) technology [140, 141], and when (or in which form) this is actually a valuable goal for law to preserve.

Step 3b: Integrate into a triadic model: recommendation and prescription.

If generative AI art models may lead to a disruption or rearticulation of widely shared notions (of the meaning, or the value) of 'creativity', then TechLaw regulatory approaches would benefit from engaging in broader (public) participation and/or (expert) debate about the intended purposes of the regulatory response to generative AI. If general-purpose generative AI creates legal challenges for the EU AI Act's application-stage, risk-focused regulatory framework, then responses would benefit from taking into consideration evolving notions of the balance of responsibility for harms throughout the AI value chain. [142, 143].

5 Evaluating the triadic model: strengths and limits

There are at least three reasons to pursue the further development, testing, and application of the triadic model as a framework for synthesizing insights from TechEthics and TechLaw. First, to technology ethicists, a triadic model foregrounds the multiple realizability of ethical interventions. Familiar routes to cope with value change are by responsive ethical initiative, or by means of technological design (altering either a technology's development process, or its artifactual features, as proposed by "ethics by design" approaches). The third route of implementing more flexible regulatory frameworks is underexplored in current TechEthics.

In response to emerging AI and associated value changes, ethical interventions via each of these routes are called-for and alignment between them is needed. Emphasis on the multiple realizability of ethical interventions facilitates a shift, away from a narrowly reactive 'problem-solving' orientation towards treating the disruptive symptoms of emerging AI technologies in diverse domains, and instead towards a general 'problem-finding' orientation towards these challenges [144]. Such problem-finding approaches to AI include strategies which do not only study how or where AI systems might create problems for existing law or ethics, to be 'solved' through law or ethics, but which instead takes stock of the ways in which AI technologies may also shape or disrupt the processes, instruments, assumptions, and even goals of existing regulatory or ethical orders.

A second benefit of the triadic approach is particularly relevant to technology lawyers. The triadic model benefits regulatory approaches by making them more tailored and resilient. Current regulatory approaches to AI tend to be either technology-centric (focused on regulating 'AI')[14]; application-centric (e.g., focused on drones; self-driving cars; facial recognition); or law-centric[15] (e.g., focusing on problems for specific doctrines such as liability or tax law) [105, 106].

While these approaches all have value, and must play a role in societal responses to the technology, they also all have shortfalls: a technology-centric approach is problematic, since as a technology AI is rather amorphous, difficult to define, and encompasses several sub-technologies whose ethical features are rather different (e.g., machine learning vs. symbolic AI). The application-centric approach, too, is

---

[14] See for instance [145]. For a detailed argument discussing the shortfalls of such an approach, see [146].

[15] What Nicolas Petit has called a 'legalistic' approach, which "consists in starting from the legal system, and proceed by drawing lists of legal fields or issues affected by AIs and robots." [147, p. 2].

problematic, since AI applications are a moving target, and at times mix together algorithmic sub-technologies within different domains. A law-centric approach has shortfalls because it is too siloed and segmented across pre-existing doctrinal lines: such a compartmentalized approach will suffer in carrying out regulatory prioritization and triage (by focusing overmuch on legally 'interesting' puzzles or edge cases); moreover, it may often result in duplication of effort at best, and "ineffective, counterproductive, or even harmful rules and policy prescriptions" [64, p. 349] at worst—with a frequent outcome regulatory fragmentation, incoherence or conflict.

A triadic approach has value here, as it supports a more holistic perspective in technology law, one which helps shift away from debates over technological exceptionalism, in order to examine new technologies (such as AI) in conjunction with the broader dynamics of social change and value change in which they are implicated. This can ground regulatory frameworks that are more resilient and efficacious.

Third, the triadic model enables more effective and meaningful triage among many technological changes, helping to identify where these may be most disruptive, and where ethical and regulatory interventions are most urgently needed. The model is specifically applicable to the second-order disruptions instigated by emerging Socially Disruptive Technologies such as AI, which—in contradistinction to first-order disruptions—frequently reveal regulatory gaps or uncertainties and provoke value changes. Indeed, part of what makes these impacts disruptive is the uncertainties they provoke and the ethical re-orientation they require. As such, the qualification "socially disruptive" can serve as a useful shorthand for sociotechnical impacts which urgently require ethical and regulatory attention, and as a decision-heuristic in situations of ethical triage under uncertainty. To ascertain whether this qualification is warranted, it does not suffice to examine AI, value change and regulation in isolation; instead, the triad should be approached in conjunction.

Of course, this is just an initial sketch. Any model picks and chooses certain elements that it deems relevant to highlight, while neglecting others. In capturing the complex ecosystem of sociotechnical change, there are different interacting domains that could be highlighted and that might, potentially, be added to the model we have sketched. Some additional nuances that the Triadic model might need to account for, and which could be the subject of fruitful future work, include: (1) the ways in which the doctrinal and legal differences between national jurisdictions affect specific TechLaw (and therefore Triadic) analyses, in ways that might not line up with broader cross-society value changes; (2) potential cases that show ambiguities in the distinction between a technology's (ethical or legal) impacts as being either first or second order.

At the same time, increasing model complexity also comes with a loss in practicability. Given the shared prescriptive orientation of the moral and legal domain, and their centrality to societal responses to technological disruption, we believe the technology triad we have sketched offers a good starting point to outline the broader ecosystem in which societal responses to Socially Disruptive Technologies can be advanced. Future work will further help to crystallize exactly what level of model complexity proves ideal for a workable and effective response.

6 Conclusion

The social impact of AI and other Socially Disruptive Technologies goes far beyond economic or industry changes, and includes changes to prevailing norms, values and legal systems that may be far-reaching. This paper has argued that an analytic framework for understanding these changes, and formulating an appropriate normative response to them, benefits from adopting a triadic model, which captures the interplay between technology, values, and regulation. We have outlined this triadic model and highlighted three particular strengths. First, to technology ethicists, a triadic model foregrounds the multiple realizability of ethical interventions and facilitates a shift from a reactive stance in the face of emerging AI, to a problem-solving and problem-finding orientation. Second, to technology lawyers, a triadic approach shifts away from debates over technological exceptionalism, and examines AI in conjunction with the broader dynamics of social change and value change in which it is implicated, grounding more tailored and resilient regulatory frameworks. Third, the triadic model enables triage among many technological changes, helping to identify where these may be most disruptive, and where ethical and regulatory interventions are most urgently needed. Applying this model facilitates an integrated and streamlined moral and legal response, which is urgently needed in the face of disruptive AI—and for the many other Socially Disruptive Technologies still waiting in the wings.

Declarations

Conflict of interest All the authors declare that they have no conflicts of interest.

References

1. Zhang, D., et al.: Artificial intelligence index report 2020. AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, Mar. 2021. Accessed: Mar. 03, (2021). [Online]. Available: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf

2. Liu, H.-Y.: The power structure of artificial intelligence. Law Innov. Technol. **10**(2), 197–229 (2018). https://doi.org/10.1080/17579961.2018.1527480

3. Kalluri, P.: Don't ask if artificial intelligence is good or fair, ask how it shifts power. Nature (2020). https://doi.org/10.1038/d41586-020-02003-2. (**Art. no. 7815**)

4. Horowitz, M. C.: Artificial intelligence, international competition, and the balance of power. Texas National Security Review, May 15, 2018. Accessed: 17 May 2018. [Online]. Available: https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/

5. Cummings, M. L., Roff, H., Cukier, K., Parakilas, J., Bryce, H.: Artificial intelligence and international affairs: disruption anticipated. Chatham House (2018). Available: https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf. Accessed 25 June 2018

6. Dafoe, A.: AI governance: a research agenda. Center for the governance of AI, future of Humanity Institute, Oxford, (2018). [Online]. Available: https://www.fhi.ox.ac.uk/govaiagenda/

7. Swierstra, T.: Nanotechnology and technomoral change. Ethics Politics, XV, 200–219, (2013)

8. Danaher, J.: Axiological futurism: the systematic study of the future of values. Futures **132**, 102780 (2021). https://doi.org/10.1016/j.futures.2021.102780

9. Köbis, N., Bonnefon, J.-F., Rahwan, I.: Bad machines corrupt good morals. Nat Hum Behav (2021). https://doi.org/10.1038/s41562-021-01128-2

10. Schuelke-Leech, B.-A.: A model for understanding the orders of magnitude of disruptive technologies. Technol. Forecast. Soc. Chang. **129**, 261–274 (2018). https://doi.org/10.1016/j.techfore.2017.09.033

11. Christensen, C. M., Raynor, M. E., McDonald, R.: What is disruptive innovation? Harvard Business Review, Dec. 01, 2015. Accessed: 13 Dec 2022. [Online]. Available: https://hbr.org/2015/12/what-is-disruptive-innovation

12. Hopster, J.: What are socially disruptive technologies?". Technol Soc **67**, 101750 (2021). https://doi.org/10.1016/j.techsoc.2021.101750

13. Gruetzemacher, R., Whittlestone, J.: The transformative potential of artificial intelligence. Futures **135**, 102884 (2022). https://doi.org/10.1016/j.futures.2021.102884

14. Grace, K., Salvatier, J., Dafoe, A., Zhang, B., Evans, O.: When will AI exceed human performance? Evidence from AI experts. Jair **62**, 729–754 (2018). https://doi.org/10.1613/jair.1.11222

15. Stein-Perlman, Z., Weinstein-Raun, B., Grace, K.: Expert survey on progress in AI. AI Impacts, Aug. 2022. Accessed: Aug. 08, 2022. [Online]. Available: https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/ (2022).

16. Wynroe K., Atkinson, D., Sevilla, J.: Literature review of transformative artificial intelligence timelines. Epoch, Jan. 17, 2023. https://epochai.org/blog/literature-review-of-transformative-artificial-intelligence-timelines (accessed 30 Jan 2023).

17. Muelhauser, L.: What open philanthropy means by 'transformative AI. Google Docs, Jun. 2019. https://docs.google.com/document/d/15siOkHQAoSBl_Pu85UgEDWfmvXFotzub31ow3A11Xvo/edit?usp=embed_facebook (accessed 16 July 2021).

18. Maas M. M.: Artificial intelligence governance under change: foundations, facets, frameworks. University of Copenhagen, Copenhagen, Denmark (2020). Available: https://drive.google.com/file/d/1vIJUAp_i41A5gc9Tb9EvO9aSuLn15ixq/view?usp=sharing. Accessed 18 Apr 2021

19. Dafoe, A.: AI governance: overview and theoretical lenses. In: The Oxford Handbook of AI Governance, Bullock, J., Chen, Y.-C., Himmelreich, J., Hudson, V. M., Korinek, A., Young, M., Zhang, B. (Eds) Oxford University Press, 2022, p. 0. Accessed: Oct. 21, 2022. [Online]. Available: https://docs.google.com/document/d/e/2PACX-1vQOQ0EBIaEu_LaJqWvdPKu8xlmrOCM6h6gq7eFHnN0Y2GPYoodQjLeilxQ8SUwnbVThXc0k_jCIsCX1/pub

20. Susskind, J.: Future politics: living together in a world transformed by tech. Oxford, United Kingdom ; New York, NY: Oxford University Press, (2018).

21. Zhang, B., Dafoe, A.: U.S. public opinion on the governance of artificial intelligence. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York NY USA: ACM, Feb. 2020, pp. 187–193. https://doi.org/10.1145/3375627.3375827.

22. Dreksler, N., et al.: Preliminary survey results: US and European publics overwhelmingly and increasingly agree that AI needs to be managed carefully," GovAI Blog, 17 Apr 2023. https://www.governance.ai/post/increasing-consensus-ai-requires-careful-management (accessed 19 Apr 2023).

23. YouGov America.: How concerned, if at all, are you about the possibility that AI will cause the end of the human race on Earth?|Daily Question," YouGov America, 03 Apr 2023. https://today.yougov.com/topics/technology/survey-results/daily/2023/04/03/ad825/3 (accessed 05 Apr 2023).

24. Ryan, M., Stahl, B.C.: Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. J. Inf. Commun. Ethics Soc. **19**(1), 61–86 (2020). https://doi.org/10.1108/JICES-12-2019-0138

25. European Commission.: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. European Commission, 21 Apr. 2021. Accessed: 07 Jul 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

26. Sheehan, M.: China's New AI Governance Initiatives Shouldn't Be Ignored. Carnegie Endowment for International Peace, 04 Jan 2022. https://carnegieendowment.org/2022/01/04/

china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127 (accessed 13 Jan 2022).

27. Schmitt, L.: Mapping global AI governance: a nascent regime in a fragmented landscape. AI Ethics (2021). https://doi.org/10.1007/s43681-021-00083-y

28. Cihon, P., Maas, M.M., Kemp, L.: Fragmentation and the future: investigating architectures for international AI governance. Global Pol. **11**(5), 545–556 (2020). https://doi.org/10.1111/1758-5899.12890

29. Garcia E. V.: Multilateralism and artificial intelligence: what role for the United Nations?," In: The global politics of artificial intelligence. Tinnirello, M., (eds) CRC Press, Boca Raton (2020), p. 18. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3779866. Accessed 14 June 2020

30. Global Partnership on Artificial Intelligence.: Joint Statement from founding members of the Global Partnership on Artificial Intelligence. 15 Jun 2020. [Online]. Available: https://www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/news/article/launch-of-the-global-partnership-on-artificial-intelligence-by-15-founding

31. UNESCO.: UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence. UNESCO, (2021). https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence (accessed 13 Jan 2022).

32. Roser, M.: Artificial intelligence is transforming our world—it is on all of us to make sure that it goes well. Our World in Data, (2022). https://ourworldindata.org/ai-impact (accessed 12 Dec 2022).

33. Nemitz, P.: Constitutional democracy and technology in the age of artificial intelligence. Phil. Trans. R. Soc. A **376**(2133), 20180089 (2018). https://doi.org/10.1098/rsta.2018.0089

34. Sparrow, R.: Killer robots. J. Appl. Philos. **24**(1), 62–77 (2007). https://doi.org/10.1111/j.1468-5930.2007.00346.x

35. Minkkinen, M., Mäntymäki, M.: Discerning between the 'Easy' and 'Hard' problems of AI governance. Forthcoming.

36. Liu, H.-Y., Maas, M., Danaher, J., Scarcella, L., Lexer, M., Rompaey, L.V.: Artificial intelligence and legal disruption: a new model for analysis. Law Innov. Technol. **12**(2), 205–258 (2020). https://doi.org/10.1080/17579961.2020.1815402

37. Moore, G.E.: Cramming more components onto integrated circuits. Electronics **38**(8), 82–85 (1965)

38. Mack, C.A.: Fifty years of Moore's law. IEEE Trans. Semicond. Manuf. **24**(2), 202–207 (2011)

39. Kott, A., Perconti, P.: Long-term forecasts of military technologies for a 20–30 year horizon: an empirical assessment of accuracy," arXiv:1807.08339 [cs], Jul. 2018, Accessed: 18, 2018. [Online]. Available: http://arxiv.org/abs/1807.08339

40. Muelhauser, L.: How feasible is long-range forecasting?. Open Philanthropy (2019). https://www.openphilanthropy.org/research/how-feasible-is-long-range-forecasting/ (accessed 25 Jun 2022).

41. Armstrong, S., Sotala, K.: How we're predicting AI—or failing to. In: Beyond Artificial Intelligence, Romportl, J., Zackova, E., Kelemen, J., (eds) in Topics in Intelligent Engineering and Informatics, vol. 9. Cham: Springer International Publishing, 2015, pp. 11–29. https://doi.org/10.1007/978-3-319-09668-1_2.

42. Gruetzemacher, R., Dorner, F., Bernaola-Alvarez, N., Giattino, C., Manheim, D.: Forecasting AI progress: a research agenda (2020). Accessed: 24 Aug 2020. [Online]. Available: http://arxiv.org/abs/2008.01848

43. MacAskill, W.: Are we living at the hinge of history? (2020) Accessed: 20 Sep 2020. [Online]. Available: https://www.academia.edu/43481026/Are_We_Living_at_the_Hinge_of_History

44. Davidson, T.: Could advanced AI drive explosive economic growth?. Open Philanthropy Project, (2021). Accessed: 10 Feb 2022. [Online]. Available: https://www.openphilanthropy.org/could-advanced-ai-drive-explosive-economic-growth

45. Roodman D.: Modeling the human trajectory. Open Philanthropy, 15 Jun 2020. https://www.openphilanthropy.org/blog/modeling-human-trajectory (accessed 31 Aug 2020).

46. Roser M.: Technology over the long run: zoom out to see how dramatically the world can change within a lifetime. Our World in Data, 06 Dec 2022. https://ourworldindata.org/technology-long-run (accessed 12 Dec 2022).

47. Davidson, T.: Semi-informative priors over AI timelines. Open Philanthropy Project (2021) Accessed: 13 Jun 2022. [Online]. Available: https://www.openphilanthropy.org/research/semi-informative-priors-over-ai-timelines/

48. Roser M.: Artificial intelligence has advanced despite having few resources dedicated to its development—now investments have increased substantially. Our World in Data (2022). https://ourworldindata.org/ai-investments (accessed 12 Dec 2022).

49. Karnofsky H.: The track record of futurists seems ... fine. Cold Takes (2022). https://www.cold-takes.com/the-track-record-of-futurists-seems-fine/ (accessed 16 Aug 2022).

50. Luu, D.: Futurist prediction methods and accuracy (2022). https://danluu.com/futurist-predictions/ (accessed 15 Sep 2022).

51. Michael, J., et al.: What do NLP researchers believe? Results of the NLP community metasurvey p. 31, (2022)

52. Zhang, B., et al.: Forecasting AI progress: evidence from a survey of machine learning researchers. arXiv (2022). https://doi.org/10.48550/arXiv.2206.04132.

53. Aguirre, A.: Will there be human-machine intelligence parity before 2040? Metaculus (2016). https://www.metaculus.com/questions/384/humanmachine-intelligence-parity-by-2040/ (accessed 18 Oct 2022).

54. Aguirre, A.: When will the first weakly general AI system be devised, tested, and publicly announced?. Metaculus (2020). https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/ (accessed 18 Oct 2022).

55. Barnett M.: When will the first general AI system be devised, tested, and publicly announced? Metaculus, (2020). https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/ (accessed 18 Oct 2022).

56. Cotra, A.: Forecasting TAI with biological anchors (Draft)," Open Philanthropy Project, (2020). [Online]. Available: https://drive.google.com/drive/folders/15ArhEPZSTYU8f012bs6ehPS6-xmhtBPP

57. Cotra, A.: Two-year update on my personal AI timelines. AI Alignment Forum (2022). https://www.alignmentforum.org/posts/AfH2oPHCApdKicM4m/two-year-update-on-my-personal-ai-timelines (accessed 03 Aug 2022).

58. Carlsmith, J.: How much computational power does it take to match the human brain? Open Philanthropy, (2020). https://www.openphilanthropy.org/brain-computation-report (accessed. 12 Dec 2020).

59. Shahar, A.: Exploring artificial intelligence futures. JAIH **2**, 169–194 (2018). https://doi.org/10.46397/JAIH.2.7

60. Karnofsky, H.: AI timelines: where the arguments, and the 'experts,' stand," cold takes, (2021). https://www.cold-takes.com/where-ai-forecasting-stands-today/ (accessed 14 Jan 2022).

61. Roser, M.: AI timelines: what do experts in artificial intelligence expect for the future?," Our World Data, (2022). https://ourworldindata.org/ai-timelines (accessed 12 Dec 2022).

62. Collingridge, D.: The social control of technology. Palgrave Macmillan, New York (1981)

63. Horowitz, M.C.: Do emerging military technologies matter for international politics? Annu. Rev. Polit. Sci. **23**(1), 385–400 (2020). https://doi.org/10.1146/annurev-polisci-050718-032725

64. Crootof, R., Ard, B.J.: Structuring Techlaw. Harvard J. Law Technol. **34**(2), 347–417 (2021)
65. Danaher, J., Sætra, H.S.: Technology and moral change: the transformation of truth and trust. Ethics Inf. Technol. **24**(3), 35 (2022). https://doi.org/10.1007/s10676-022-09661-y
66. Swierstra, T., Stemerding, D., Boenink, M.: Exploring techno-moral change: the case of the obesitypill. In: Sollie, P., Duwell, M. (eds.) Evaluating new technologies, pp. 119–138. Springer, Dordrecht (2009)
67. Kudina, O.: The technological mediation of morality: value dynamism, and the complex interaction between ethics and technology. [PhD Thesis - Research UT, graduation UT, University of Twente]. University of Twente (2019). https://doi.org/10.3990/1.9789036547444
68. Nickel, P. J., Kudina, O., van de Poel, I.: Moral uncertainty in technomoral change: bridging the explanatory gap. Perspect. Sci. (2022) Accessed: 18 Nov 2021. [Online]. Available: https://philpapers.org/archive/NICMUI-2.pdf
69. Friedman, B., Hendry, D.G.: Value sensitive design: shaping technology with moral imagination. MIT Press, Cambridge (2019)
70. Umbrello, S., van de Poel, I.: Mapping value sensitive design onto AI for social good principles. AI Ethics **1**(3), 283–296 (2021). https://doi.org/10.1007/s43681-021-00038-3
71. Project SHERPA.: Ethics by design. Project SHERPA. https://www.project-sherpa.eu/ethics-by-design/ (accessed 14 Jan 2022).
72. Winner, L.: Do artifacts have politics? Daedalus **109**(1), 121–136 (1980)
73. Brey, P.A.E.: Anticipatory ethics for emerging technologies. NanoEthics **6**(1), 1–13 (2012). https://doi.org/10.1007/s11569-012-0141-7
74. Maas, M.M.: International law does not compute: artificial intelligence and the development, displacement or destruction of the global legal order. Melb. J. Int. Law **20**(1), 29–56 (2019)
75. Rotolo, D., Hicks, D., Martin, B.: What is an emerging technology? Res. Policy **44**(10), 1827–1843 (2015)
76. Bennett Moses, L.: Why have a theory of law and technological change? Minnesota J. Law Sci. Technol. **8**(2), 589–606 (2007)
77. Abbott, K.: Introduction: the challenges of oversight for emerging technologies. In: Innovative Governance Models for Emerging Technologies, Edward Elgar Publishing, 2013. Accessed: 24 Jul 2018. [Online]. Available: https://www.elgaronline.com/view/9781782545637.00006.xml
78. Picker, C.B.: A view from 40,000 feet: international law and the invisible hand of technology. Cardozo Law Rev. **23**, 151–219 (2001)
79. Maas, M.M.: Innovation-proof governance for military AI? How I learned to stop worrying and love the bot. J. Int. Humanitarian Legal Studies **10**(1), 129–157 (2019). https://doi.org/10.1163/18781527-01001006
80. Crootof, R.: Jurisprudential space junk: treaties and new technologies. In: Resolving conflicts in the law, Giorgetti, C., Klein, N (eds) 2019, pp. 106–129. Accessed:. 15 Mar 2019. [Online]. Available: https://brill.com/view/book/edcoll/9789004316539/BP000015.xml
81. Eichensehr, K.E.: Cyberwar & international law step zero. Texas Int. Law J. **50**(2), 357–380 (2015)
82. Israel, B.: Treaty stasis. AJIL Unbound, vol. 108, pp. 63–69, ed 2014, doi: https://doi.org/10.1017/S2398772300001860.
83. Marchant, G. E.: The growing gap between emerging technologies and the law. In: The growing gap between emerging technologies and legal-ethical oversight: the pacing problem, Marchant, G. E., Allenby, B. R., Herkert, J. R. (eds) in The International Library of Ethics, Law and Technology. Dordrecht: Springer Netherlands, 2011, pp. 19–33. https://doi.org/10.1007/978-94-007-1356-7_2.
84. Allenby, B. R.: The dynamics of emerging technology systems. In: The growing gap between emerging technologies and legal-ethical oversight: the pacing problem. Marchant, G. E., Herkert, J. R. (eds) in The International Library of Ethics, Law and Technology. Springer Netherlands, 2011. Accessed: 15 May 2018. [Online]. Available: //www.springer.com/gp/book/9789400713550
85. Bennett Moses, L.: Agents of change: how the law 'copes' with technological change. Griffith Law Rev. **20**(4), 763–794 (2011). https://doi.org/10.1080/10383441.2011.10854720
86. Ard, B., Crootof, R.: The case for 'technology law. Nebraska Governance & Technology Center, 16 Dec 2020. https://ngtc.unl.edu/blog/case-for-technology-law (accessed 16 Mar 2021).
87. Friedman, D.D.: Does technology require new law? Public Policy **71**, 16 (2001)
88. Bennett Moses, L.: Regulating in the face of sociotechnical change. In: The Oxford Handbook of Law, Regulation, and Technology, Brownsword, R., Scotford, E., Yeung, K. (eds), pp. 573–596 (2017). Accessed: 13 May 2017. [Online]. Available: http://www.oxfordhandbooks.com/view/https://doi.org/10.1093/oxfordhb/9780199680832.001.0001/oxfordhb-9780199680832-e-49
89. Jones, M.L.: Does technology drive law? The dilemma of technological exceptionalism in cyberlaw. SSRN J (2018). https://doi.org/10.2139/ssrn.2981855
90. Forsberg, E.-M.: ELSA and RRI—editorial. Life Sci Soc Policy **11**, 2 (2015). https://doi.org/10.1186/s40504-014-0021-8
91. von Schomberg R.: Prospects for technology assessment in a framework of responsible research and innovation. In: Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden, Dusseldorp, M., Beecroft, R. (eds) Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 39–61. (2012). https://doi.org/10.1007/978-3-531-93468-6_2.
92. Owen, R., Macnaghten, P., Stilgoe, J.: Responsible research and innovation: from science in society to science for society, with society. Sci Public Policy **39**(6), 751–760 (2012). https://doi.org/10.1093/scipol/scs093
93. Floridi, L., et al.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind. Mach. **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
94. Martin, J.L.: Structuring the sexual revolution. Theor. Soc. **25**(1), 105–151 (1996). https://doi.org/10.1007/BF00140760
95. Hopster, J.K.G., et al.: Pistols, pills, pork and ploughs: the structure of technomoral revolutions. Inquiry (2022). https://doi.org/10.1080/0020174X.2022.2090434
96. Danaher, J., Hopster, J.: The normative significance of future moral revolutions. Futures (2022). https://doi.org/10.1016/j.futures.2022.103046
97. Floridi, L., Strait, A.: Ethical foresight analysis: what it is and why it is needed? Mind. Mach. (2020). https://doi.org/10.1007/s11023-020-09521-y
98. Keulartz, J., Schermer, M., Korthals, M., Swierstra, T.: Ethics in technological culture: a programmatic proposal for a pragmatist approach. Sci. Technol. Human Values **29**(1), 3–29 (2004). https://doi.org/10.1177/0162243903259188
99. van der Burg, W.: Dynamic ethics. J. Value Inq. **37**(1), 13–34 (2003). https://doi.org/10.1023/A:1024009125065
100. Brownsword, R., Scotford, E., Yeung, K.: Law, regulation, and technology: the field, frame, and focal questions. In: The Oxford Handbook of Law, Regulation and Technology, Brownsword, R., Scotford, E., Yeung, K. (eds) Oxford University Press, (2017). https://doi.org/10.1093/oxfordhb/9780199680832.013.1.
101. Easterbrook, F.H.: Cyberspace and the law of the horse. The University of Chicago Legal Forum **207**, 11 (1996)

102. Lessig, L.: The law of the horse: what cyberlaw might teach. Harv. Law Rev. **113**(2), 501 (1999). https://doi.org/10.2307/1342331

103. Calo, R.: Robotics and the lessons of cyberlaw. Calif. L. Rev. **103**, 513–564 (2015)

104. Balkin, J.M.: The path of robotics law. Calif. Law Rev. Circuit **6**, 17 (2015)

105. Petit, N., De Cooman, J.: Models of law and regulation for AI. Social Science Research Network, EUI Working Paper RSCAS 2020/63 ID 3706771, (2020). https://doi.org/10.2139/ssrn.3706771.

106. Maas, M. M.: Aligning AI regulation to sociotechnical change. In: The Oxford Handbook of AI Governance, (2022). https://doi.org/10.1093/oxfordhb/9780197579329.013.22.

107. Aizenberg, E., van den Hoven, J.: Designing for human rights in AI. Big Data Soc. **7**(2), 2053951720949566 (2020). https://doi.org/10.1177/2053951720949566

108. Smuha, N. A.: Beyond the individual: governing AI's societal harm. Internet Policy Review, **10**, (3), (2021). Accessed: 12 Oct 2021. [Online]. Available: https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm

109. Helbing, D., et al.: Will democracy survive big data and artificial intelligence?, Sci Am (2017) Accessed: 29 May 2017. [Online]. Available: https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

110. Chesney, R., Citron, D.K.: Deep fakes: a looming challenge for privacy, democracy, and national security. Calif. Law Rev. **107**, 1753–1820 (2019)

111. Brownsword, R.: From Erewhon to AlphaGo: for the sake of human dignity, should we destroy the machines? Law Innov. Technol. **9**(1), 117–153 (2017). https://doi.org/10.1080/17579961.2017.1303927

112. Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: can language models be too big? 🦜," In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, in FAccT '21. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 610–623. https://doi.org/10.1145/3442188.3445922.

113. Smith, B.W.: New technologies and old treaties. Am. J. Int. Law **114**, 152–157 (2020). https://doi.org/10.1017/aju.2020.28

114. Liu, H.-W., Lin, C.-F.: Artificial intelligence and global trade governance: a pluralist agenda. Harvard Int. Law J. **61**(2), (2020), Accessed: Sep. 26, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3675505

115. Docherty, B.: The need for and elements of a new treaty on fully autonomous weapons. Human Rights Watch, (2020). https://www.hrw.org/news/2020/06/01/need-and-elements-new-treaty-fully-autonomous-weapons (accessed 03 Jun 2020).

116. King, T.C., Aggarwal, N., Taddeo, M., Floridi, L.: Artificial intelligence crime: an interdisciplinary analysis of foreseeable threats and solutions. Sci Eng Ethics (2019). https://doi.org/10.1007/s11948-018-00081-0

117. Hayward, K.J., Maas, M.M.: Artificial intelligence and crime: a primer for criminologists. Crime Media Cult. **17**(2), 209–233 (2020). https://doi.org/10.1177/1741659020917434

118. Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M.: Defining organizational AI governance. AI Ethics **2**(4), 603–609 (2022). https://doi.org/10.1007/s43681-022-00143-x

119. Crootof, R., Kaminski, M. E., Price II, W. N.: Humans in the loop. Soc. Sci. Res. Netw., Rochester, NY, SSRN Scholarly Paper ID 4066781, (2022). https://doi.org/10.2139/ssrn.4066781.

120. Dafoe, A.: On technological determinism: a typology, scope conditions, and a mechanism. Sci. Technol. Human Values **40**(6), 1047–1076 (2015). https://doi.org/10.1177/0162243915579283

121. Sheppard, B.: Warming up to inscrutability: How technology could challenge our concept of law. Univ. Toronto Law J. **68**(supplement 1), 36–62 (2018). https://doi.org/10.3138/utlj.2017-0053

122. Hopster, J.: The ethics of disruptive technologies: towards a general framework. In: new trends in disruptive technologies, tech ethics and artificial intelligence, de Paz Santana, J. F., de la Iglesia, D. H., López Rivero, A. J., (eds) in Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2022, pp. 133–144. https://doi.org/10.1007/978-3-030-87687-6_14.

123. Nissenbaum, H.: Contextual integrity up and down the data food chain. Theoretical Inquiries in Law, **20**(1), Art. no. 1, (2019). Accessed:. 14 Jan 2022. [Online]. Available: https://www7.tau.ac.il/ojs/index.php/til/article/view/1614

124. Bommasani, R. et al.: On the opportunities and risks of foundation models. arXiv (2022). https://doi.org/10.48550/arXiv.2108.07258.

125. Schneider, J.: Foundation models in brief: a historical, sociotechnical focus. arXiv (2022). Accessed: 08 Jan 2023. [Online]. Available: http://arxiv.org/abs/2212.08967

126. Gutierrez, C. I., Aguirre, A., Uuk, R., Boine, C. C., Franklin, M.: A proposal for a definition of general purpose artificial intelligence systems. (2022). https://doi.org/10.2139/ssrn.4238951.

127. Hacker, P., Engel, A., Mauer, M.: Regulating ChatGPT and other large generative AI models (2023). Available: https://europeannewschool.eu/images/chairs/hacker/Hacker_Engel_Mauer_2023_Regulating_ChatGPT_Feb07.pdf. Accessed 10 Feb 2023

128. Carlson, A.: Regulating ChatGPT and other language models: a need for balance. Astrafizik, (2022). https://astrafizik.com/eng/tech/regulating-chatgpt-and-other-language-models-a-need-for-balance/ (accessed 20 Jan 2023).

129. Weidinger, L. et al.: Taxonomy of risks posed by language models," In 2022 ACM conference on fairness, accountability, and transparency, Seoul Republic of Korea: ACM, pp. 214–229 (2022). https://doi.org/10.1145/3531146.3533088.

130. J. Okerlund et al.: What's in the Chatterbox? Large language models, why they matter, and what we should do about them. Ford School of Public Policy, University of Michigan, (2022). Available: https://stpp.fordschool.umich.edu/research/research-report/whats-in-the-chatterbox. Accessed 20 Jan 2023

131. Future of Life Institute.: Pause giant AI experiments: an open letter. Future of Life Institute, (2023). https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (accessed 30 Mar 2023).

132. Mukherjee, S., Pollina, E., More, R.: "Italy's ChatGPT ban attracts EU privacy regulators," Reuters, (2023). Accessed: 04 Apr 2023. [Online]. Available: https://www.reuters.com/technology/germany-principle-could-block-chat-gpt-if-needed-data-protection-chief-2023-04-03/

133. Vincent, J.: The lawsuit that could rewrite the rules of AI copyright. The Verge, Nov. 08, (2022). https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data (accessed 20 Jan 2023).

134. "How will China's Generative AI Regulations Shape the Future? A DigiChina Forum," DigiChina, Apr. 19, 2023. https://digichina.stanford.edu/work/how-will-chinas-generative-ai-regulations-shape-the-future-a-digichina-forum/ (accessed 05 May 2023).

135. Porsdam Mann, S. et al.: Generative AI entails a credit–blame asymmetry. Nat Mach Intell, pp. 1–4, (2023). https://doi.org/10.1038/s42256-023-00653-1.

136. Volpicelli, G.: ChatGPT broke the EU plan to regulate AI. POLITICO, (2023). https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/ (accessed 20 Apr 2023).

137. Helberger, N., Diakopoulos, N.: ChatGPT and the AI Act. Internet Policy Review, **12**(1) (2023), Accessed: 21 Feb 2023. [Online]. Available: https://policyreview.info/essay/chatgpt-and-ai-act

138. AI Now Institute, "General Purpose AI Poses Serious Risks, Should Not Be Excluded From the EU's AI Act | Policy Brief," AI Now Institute, (2023). https://ainowinstitute.org/publication/gpai-is-high-risk-should-not-be-excluded-from-eu-ai-act (accessed 16 Apr 2023).

139. Prettner, C.: FLI position paper on AI liability. Future of Life Institute, 2022. Accessed: 28, Apr 2023. [Online]. Available: https://futureoflife.org/wp-content/uploads/2022/11/FLI_AI_Liability_Position_Paper.pdf

140. Seger, E.: What do we mean when we talk about 'AI democratisation'?," GovAI Blog (2023). https://www.governance.ai/post/what-do-we-mean-when-we-talk-about-ai-democratisation (accessed 10 Feb 2023).

141. Seger, E., Ovadya, A., Garfinkel, B., Siddarth, D., Dafoe, A.: Democratising AI: multiple meanings, goals, and methods. arXiv, (2023). https://doi.org/10.48550/arXiv.2303.12642.

142. Engler, A., Renda, A.: Reconciling the AI value chain with the EU's artificial intelligence Act. CEPS, (2022). Accessed: 28 Apr 2023. [Online]. Available: https://www.ceps.eu/ceps-publications/reconciling-the-ai-value-chain-with-the-eus-artificial-intelligence-act/

143. Küspert, S., Moës, N., Dunlop, C.: The value chain of general-purpose AI," Ada Lovelace Institute, (2023). https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/ (accessed 28 Apr 2023).

144. Liu, H.-Y., Maas, M.M.: 'Solving for X?' Towards a problem-finding framework to ground long-term governance strategies for artificial intelligence. Futures **126**, 22 (2021). https://doi.org/10.1016/j.futures.2020.102672

145. Turner, J.: Robot rules: regulating artificial intelligence. Springer Berlin, Heidelberg (2018)

146. Schuett, J.: Defining the scope of AI regulations. Law, Innovation and Technology, **15**(1) forthcoming (2023). Available: https://doi.org/10.1080/17579961.2023.2184135. Accessed 6 Mar 2023

147. Petit, N.: Law and regulation of artificial intelligence and robots—conceptual framework and normative implications. Soc. Sci. Res. Netw. Rochester, NY, SSRN Scholarly Paper ID 2931339, (2017). Accessed: 11 May 2020. [Online]. Available: https://papers.ssrn.com/abstract=2931339