Henrik Skaug Sætra *is an Associate Professor at Østfold University College. He is a political scientist with a broad and interdisciplinary approach to issues of ethics and the individual and societal implications of technology, environmental ethics, and game theory.*

Eduard Fosch-Villaronga *is Assistant Professor at the eLaw Center for Law and Digital Technology at Leiden University, where he investigates the legal and regulatory aspects of robot and AI technologies, with a special focus on healthcare.*

Research in AI has Implications for Society: How do we Respond?

THE BALANCE OF POWER BETWEEN SCIENCE, ETHICS, AND POLITICS

KEYWORDS:
ARTIFICIAL INTELLIGENCE, ETHICAL, LEGAL, AND SOCIETAL IMPLICATIONS, SCIENCE, ETHICS, POLITICS
DOI:
https://doi.org/10.5771/2747-5174-2021-1-60

ABSTRACT:
Artificial intelligence (AI) offers previously unimaginable possibilities, solving problems faster and more creatively than before, representing and inviting hope and change, but also fear and resistance. Unfortunately, while the pace of technology development and application dramatically accelerates, the understanding of its implications does not follow suit. Moreover, while mechanisms to anticipate, control, and steer AI development to prevent adverse consequences seem necessary, the current power dynamics on which society should frame such development is causing much confusion. In this article we ask whether AI advances should be restricted, modified, or adjusted based on their potential legal, ethical, societal consequences. We examine four possible arguments in favor of subjecting scientific activity to stricter ethical and political control and critically analyze them in light of the perspective that science, ethics, and politics should strive for a division of labor and balance of power rather than a conflation. We argue that the domains of science, ethics, and politics should not conflate if we are to retain the ability to adequately assess the adequate course of action in light of AI's implications. We do so because such conflation could lead to uncertain and questionable outcomes, such as politicized science or ethics washing, ethics constrained by corporate or scientific interests, insufficient regulation, and political activity due to a misplaced belief in industry self-regulation. As such, we argue that the different functions of science, ethics, and politics must be respected to ensure AI development serves the interests of society.

AUTHORS: Henrik Skaug Sætra & Eduard Fosch-Villaronga

Artificial intelligence (AI) offers possibilities previously unimaginable, solving problems in new and innovative ways. AI represents and invites hope and change, but also fear and resistance. Unfortunately, while the pace of technology development and their applied uses for research dramatically accelerate, the understanding of its implications does not follow in parallel. Moreover, control mechanisms to restrict or redirect certain technological advances in light of potential adverse consequences to society seem inadequate, and some argue that ethics should be embedded in research and business or that business can or should self-regulate. Although frameworks like Responsible Research and Innovation (RRI) promote reflection upon the implications of technology outcomes and foster the incorporation of such considerations into the research and design processes, they say little about how to proceed when such implications are perceived to be excessively adverse. For example, should autonomous weapon systems exist (Sparrow, 2007)? Should sex robots, or social robots in general, be further developed (Fosch-Villaronga & Poulsen, 2020; Levy, 2009; Sætra, 2020a, 2020b, 2021b; Sullins, 2012)? Should scientists explore the use of life-like child robots to treat pedophilia (Danaher, 2019)? Should philosophers research whether robots deserve rights or personhood (Birhane & van Dijk, 2020; Coeckelbergh, 2010; Gellers, 2020; Gunkel, 2018)? And not least, should algorithmic decision-making and facial recognition technology be developed, or deployed, despite their potential for discriminatory practices (Coalition for Critical Technology, 2020)?

Human existence is a web of interrelated processes in which our choices and actions affect most other parts of this web. This intertwined reality implies that scientific activity potentially always has societal implications (Næss, 1989). This link between science and societal consequences leads to debates about the scientists' role and responsibility and whether and how science should be promoted, allowed, controlled, restricted, or banned (ChoGlueck, 2018; Kitcher, 2003; Kourany, 2003). These debates relate both to the fact that values are inevitably present in science, but also to the power of science and the need to control it in order to achieve certain social goals. To highlight some of the reasons to restrict research on AI and new technology, and subject it to ethical or political control, we examine four logical arguments in favor of restricting scientific activity and critically analyze these in light of the ideal of non-conflation of science, ethics, and politics.

We ask if the domains of science, ethics, and politics should perhaps not be conflated if we are to retain the ability to adequately assess the course of action in light of AI's implications. This is because such conflation could lead to uncertain and questionable outcomes, such as politicized science or ethics washing, ethics constrained by corporate or scientific interests, or insufficient regulation and political activity due to a misplaced belief in industry self-regulation (Redding, 2013; Sætra, Forthcoming; Walker & Wan, 2012). While the different functions of science, ethics, and politics must be respected, the work in each domain should be informed by the others to foster interdisciplinary collaboration or the effective implementation of (privacy, ethics)-by-design principles or well-informed regulation.

For instance, robot-oriented regulations framing AI development may be premature, misguided, or even dangerous because these technologies are at an early stage (Brundage & Bryson, 2016). This relates to the so-called responsibility gap generated by AI (Matthias, 2004). Since AI is unpredictable, and uses error as a method, some argue that developers and designers cannot be held accountable for the actions of such machines, and that not accounting for this will stifle innovation and lead to adverse societal effects (Gunkel, 2017; Matthias, 2004). Others, however, argue that such a gap is illusory, and that relieving human beings of their responsibility creates a moral hazard involving great societal risk (Sætra, 2021a). Sætra (2021a) argues that novel situations are indeed created by new developments in AI, but that these only highlight and emphasize the need for active and robust regulation. Innovation is important, and misconceived regulation, i.e., the belief that robots could dehumanize caring practices (European Parliament 2017, 2019), could hinder the development of assistive robotics, such as feeding-robots that allow for increased privacy during mealtime (Herlant, 2018), robots for the blind that improve users' autonomy and help assistance-dogs to avoid welfare-threatening punishment-based training (Bremhorst et al., 2018; Zardiashvili & Fosch-Villaronga, 2020).

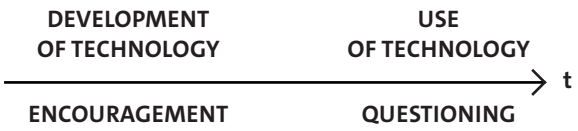
In this paper we argue that a proper division of labor etween the three domains must be found to ensure that science can push the boundaries of our existing knowledge while remaining safe; ethics can be free from the interests of science, business, and politics; and politics can perform its necessary role as the final arbiter on how and when scientific developments can be pursued and deployed in society.

QUESTIONING THE USE AND DEVELOPMENT OF AI SCIENTIFIC ACTIVITY

Some seem inclined to take technology as a given and assume that it is conducive to (positive) progress (Floridi et al., 2018). The authors propose the idea that the contemporary debate no longer focuses on whether AI impacts society or not, but whether this impact is going to be positive or negative, and to what extent. They present opportunities and risks associated with AI, compile principles supporting the adoption of AI, and put forward recommendations serving as the basis for establishing a Good AI Society. A close

reading of the article reveals the assumption that AI uptake is inevitable. In Floridi et al.'s (2018) words, „AI can be used to foster human nature and its potentialities, thus creating opportunities; underused, thus creating opportunity costs, or overused and misused, thus creating risks."

Similarly, the European project INBOTS claims that what we need is „inclusive robotics for a better society."[1] UNESCO also states, „'given the complexity of contemporary global challenges, such as the sustainable consumption of resources and climate change adaptation, supporting and investing in engineering education is essential for the improvement of societies."[2] These statements form a general pattern that encourages technology development while questioning that same technology's various applications (Fosch-Villaronga, 2019). This raises the issue of whether technology development should be restricted and put into question, for example, for the sake of preventing killer robots or sex robots that may lead to morally undesirable consequences. [3]



Encouragement and questioning of the development and use of a technology (Fosch-Villaronga, 2019).

The general topic we examine is not new. Historically, this discussion has focused on the link between blue-sky research and potentially malign applications of its findings down the road. Innovations in physics and the resulting atom bomb are classic examples (Collingridge, 1980; Merton, 1973). In this sense, a question arises: should scientific ctivities – the development of new technologies – be subject to restrictions due to the research's potential ethical and political implications? Or, rather, should we encourage any development of technology, and only intervene when it is time to deploy it beyond the controlled domain of science? While it is necessary and uncontroversial that science should be conducted ethically and under robust parameters, we ask whether even science conducted according to ethical research standards should be restricted due to potentially problematic usage of the outcomes of such science.

AI raises new questions, or at the very least makes old problems more pressing. Given the huge investment in AI as the solution to so many contemporary challenges, and the fact that technology is both a filter and an agent in determining how individuals see the world, it is of great importance to have a renewed focus on these old questions (Verbeek, 2015). Collingridge (1980) discusses technology's social control and points to its inevitable „unanticipated social consequences," which can be negative or positive (Boocher, 2012). Collingridge's dilemma refers to the hypothesis that technologies can be shaped in their infancies, but their implications are not well understood. When technologies mature and their impacts become apparent, they have become difficult to control. This insight, along with Latour's (1999) idea of black-boxing and the opinion that „scientific and technical work is made invisible by its own success," is one source of the call for tighter control of technology and thus the science behind it.

We argue for open and free science, but this in no way implies that we should simultaneously allow the uncritical deployment of new technologies (Johnston, 2018). There is no necessary contradiction between scientific freedom and relatively tight political regulation of using and developing new technologies (Sætra, 2021a). While we return to the role of politics and the law towards the end of the article, for now, it suffices to note these initial premises and mention that we will not go into detail concerning the practical issues of regulation and law.

UNQUESTIONED SCIENCE

Faith in science has traditionally been more universal and unquestioned (Merton, 1942). There are countless examples of how technology has been proposed as the solution to challenging engineering practice, government policy failures, or modern consumerism outcomes, showing how technological fixes have cultural, ethical, and political implications (Bauman, 2013; Johnston, 2018). Some have even argued that science and engineering is a form of master discipline that should even take the place of politics in technocratic societies (Meynaud, 1969; Sætra, 2020c). With the constant progress and achievement, scientists have regarded themselves as independent of society and considered science a self-validating enterprise that was "in society but not of it" (Merton, 1942). In a similar vein, tech companies today write and define reality, the meaning of societally-relevant concepts, such as privacy, and determine what is valid, appropriate, and toxic, with often disastrous consequences (Buolamwini, & Gebru, 2018; Gomes, Antonialli, & Dias-Oliva, 2019; Poulsen, Fosch-Villaronga, Søraa, 2020). Big Tech also controls modern media, and social media provide a means to reach most aspects of modern individuals (Foer, 2017; Zuboff, 2019). More worrisome is the continuous use of inferential analytics methods guessing user characteristics and preferences to support ulterior decision-making processes that significantly affect people in various ways (Nisevic et al., 2021). Big Data in combination with the techniques of nudging, for example, raises concerns about manipulation and an increasing lack of autonomy and liberty (Sætra, 2019b; Yeung, 2017).

However, scientists are an integral part of society, and this also comes with corresponding obligations and interests, as many of our actions have a wide array of potentially problematic societal impacts that are not readily identifiable for us (Merton, 1942). In our time, data scientists and the AI community are entangled in debates about structural racism, biased data, and the discriminatory effects of algorithms (Noble, 2018; Fosch-Villaronga et al., 2020),

highlighting the importance of reinvigorating and reestablishing an ethos of science, which purportedly constitutes a set of norms and values that ideally guide and unite scientists. Skeptics stress that there is some doubt about whether such norms actually bind scientists, especially since there is no consensus on what constitutes such an ethos. Moreover, the contents of such an ethos are norms - not laws - and as such, they do not lead to binding consequences and often entail a mere apology to society, e.g., when Google apologized for its technology labeling dark-skinned people as gorillas (Grush, 2015), or a quick fix, e.g., when three years later Google removed gorillas from their image-labeling technology (Vincent, 2018).

The Mertonian norms have been influential in governing science for decades, and we will use two of Merton's (1942) norms as arguments against the conflation of science and the other domains: universalism and organized[4] skepticism. These two are selected because they are particularly relevant for determining the proper division of labor between scientists, ethicists, and politicians. These norms provide a strong defense for comparatively free science – a freedom which is increasingly coming under attack in the context of research on AI, as seen, for example, in the calls for the ban of research on facial recognition technology (Coalition for Critical Technology, 2020) or killer and sex robots. Universalism refers to the universal nature of science and how all claims should be evaluated without considering their protagonist's „race, nationality, religion, class, and personal qualities" (Merton, 1942). There is a strong belief in the possibility of objectivity, and this „precludes particularism" (Merton, 1942). Of relevance to current debates, we may note that while „the chauvinist" may purge undesirable persons and facts from history, „their formulations remain indispensable to science and technology" (Merton, 1942). This is also related to other's call for diversity and a plurality of voices, theories, and ideas in science (Feyerabend, 1970, 1993). Some suggest that even "the ramblings of madmen" are valuable input in the marketplace of ideas, an idea often attributed to John Stuart Mill (1985).

One consequence of universalism is that it creates an imperative for openness for talents in science. Recruitment and access to the world of science must be open to all, regardless of the previously noted characteristics, such as race and nationality, and it is thus a radically inclusive ideal (Merton, 1942). Merton (1942) also connects this norm to the ethos of democracy and states that achieving both requires „the progressive elimination of restraints upon the exercise and development of socially valued capacities." This quickly turns into a justification of politics and regulation that is necessary for any society in which the free market permits inequalities that are not based on differences in capacity (Merton, 1942). The first norm, then, suggests that research is independent of its protagonists. However, it also points to the need for political intervention if the world of science is not equally available to all with equal capabilities. Universalism is thus a reason to ensure that everyone has equal access to the life of science. While historically unequal access to science constitutes a breach of universalism, the norm itself does not imply that we must discard the past science.

The second norm is organized skepticism – a "methodological and institutional" mandate (Merton, 1942). The disinterested and detached scientist "suspends judgment" in lieu of "empirical and logical criteria," and this approach often leads to conflicts between scientists and others (Merton, 1942). Nothing is sacred for Merton's scientist; nothing demands uncritical respect. In short, everything is fair game to the scientist, and even the most fundamental truths and values of society are never taken as axioms assumed to be true. The scientist challenges everything, and by doing so, they either discover that established truths were unfounded or contribute to a greater understanding of why the truths are indeed important (Mill, 1985).
This norm undermines the arguments in favor of conflation of domains as it explicitly calls for the suspension of judgment and emphatically demands the separation of the domains of science, ethics, and politics. We see these three domains as different functions of the system of science-in-society, where one is the producer of new knowledge, the next has the role of evaluating the implications of this knowledge, and the third (politics) sets the boundaries for the world of science and regulates how new scientific progress can be applied outside the domain of controlled science. It is akin to the political division of power, where we ideally have different powers that a) perform different tasks and specialize at these, and b) balance and counteract each other.

CONFLATION OR NOT? THAT'S THE QUESTION

The critical question we attempt to answer is whether new developments in AI warrant the conflation of the domains of science, ethics, and politics. By relying on the norms just discussed, for example, one might argue that a scientist's morality should not influence our evaluation of their work. Similarly, we should also be open to discussing, pursuing, and evaluating immoral ideas. The domains can never be fully separated, and while we encourage interdisciplinarity and collaboration across these boundaries, the functions of the three domains must remain intact. Fully conflating these disciplines, we argue, could lead to critical societal risks, as shown in Figure 2.

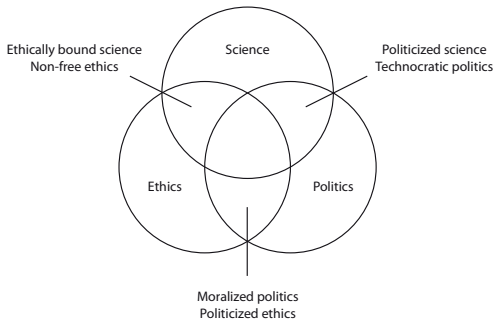


FIG. 2: THE DANGERS OF CONFLATION.

When science and ethics is conflated, science suffers by being restricted by ethics in its infancy, and innovation is stifled. Simultaneously, ethics conducted within the domain of science suffers from being exposed to the interests of cience and, in the case of AI, very often the interests of business and Big Tech (Sætra, Forthcoming). When science and politics conflate, science gains undue access to the political domain, creating the dangers of undemocratic and technocratic developments. Simultaneously, science becomes politicized and unfree. When ethics and politics conflate and overlap, something we focus relatively less on in this article, politics become moralized and ethics politicized. When all domains conflate, we get all the potential adverse effects. The scientific basis of progress and innovation, broadly understood, is undermined, as we argue that progress is based on the serendipitous and unpredictable nature of free and disinterested science.

In the context of AI, the High-Level Expert Group (HLEG) on AI released ethical guidelines on trustworthy AI. While the rights-based approach of the HLEG was remarkable, the HLEG should have avoided the conflation of fundamental rights and ethics in the concept of "ethical purpose." According to some researchers at Tilburg University, "law and ethics are two separate domains that need to be clearly distinguished with regard to their rationale and function" (Noorman et al., 2019). According to them, not doing so runs the risks of obscuring and down-playing the central role of law in the governance of the design, deployment, and use of AI in favor of ethics as a mere reinterpretation of "industry self-regulation." The role of ethics is repeatedly questioned in European legal reports, where it is stated that "there is much confusion as to what should be the relation between such ethical frameworks or guidance and fundamental rights safeguards. Such confusion, if further sustained, might be, in the end, detrimental to the protection of fundamental rights to the extent that it can divert attention from the necessity of safeguarding certain legal obligations" (González-Fuster, 2020).

In this sense, a conflation leads to unfortunate consequences, and much uncertainty within all domains. Science is about exploration and the quest for knowledge. While it is difficult to anticipate the potential adverse consequences these advancements have on society, there is no apparent reason to restrict them if such research is conducted in a controlled environment where undesirable impacts on society are prevented (Fosch-Villaronga & Heldeweg, 2018). Ethics involves the evaluation of good and bad – right or wrong. The role of ethics is to evaluate the implications of political ideas and scientific advances and inform politics about these implications. It is politics that then mandates – ethicists cannot mandate or enforce. Ethics also allows for a legitimate restriction of scientific activity, enforced by political regulation, or evaluates the consequences of new findings and technologies. When science is free, it can enable the construction of autonomous weapons, for instance. The role of ethics is to develop the criteria for evaluating the creation and potential application of such technologies. Politics uphold social order and form the basis for creating a good society, and whatever that may entail is assumed to be up to the public in this given society. While moral philosophers may inform the debate of what is considered acceptable, ultimately, those in power define what is considered suitable for any given society (Sætra, 2021a). In a democracy, the citizens will elect representatives that determine what is right on their behalf, then regulate and enforce it.

WHY WE SHOULD *NOT* RESTRICT BASIC DEVELOPMENT IN AI

To examine how the conflation of science, ethics, and politics manifests itself in current AI debates, we introduce four types of logical arguments involving some form of conflation of the domains. After each example, the premises involving conflation are critically appraised and replaced by other premises. Regarding whether a specific controversial scientific activity should be allowed, we examinefour potential responses:

1. Restrict because it is certainly unethical
2. Restrict because it could be used unethically
3. Restrict because it distracts from something more important
4. Restrict because person X has a particular characteristic

A: "RESTRICT BECAUSE IT IS CERTAINLY UNETHICAL"

A scientific approach should not be pursued or published because it has direct and known negative consequences, including being offensive to particular groups, and is considered morally problematic. The problem is not whether the research is conducted ethically, without informed consent or the respectful treatment of human or animal subjects, etc., but because ethicists find the implications of the scientific activity problematic.

THE ARGUMENT

p_1 The implications of theory/ approach Y is morally problematic
p_2 Morally problems science should not be conducted
p_3 The scientific activity X is based on theory/approach Y
q the scientific activity X should be suppressed

An example involves the controversy surrounding a paper about facial recognition to predict criminality (Hao, 2020). The article was due to be published by Springer Nature. However, opponents of this particular use of such technology formed a coalition to stop the publication (Coalition for Critical Technology, 2020). An open letter to Springer was written and signed by the coalition members and sent to Springer Nature urging them to "publicly rescind the offer for the publication of this specific study," "issue a statement condemning the use of criminal justice statistics to predict criminality," and "refrain from publishing similar studies in the future" (Coalition for Critical Technology, 2020). The reasons stated were a) that the technologies discussed are based on "unsound scientific premises, research, and methods," and b) that it is vital to keep such technologies from being developed as they are not neutral. Another argument was that governments use them for "depoliticizing state violence and reasserting the legitimacy of the carceral state" (Coalition for Critical Technology, 2020).

Springer Nature subsequently stated that they would not publish the article, and the authors of the article requested that the university remove the press release about the article in question (Hao, 2020; Harrisburg University, 2020). The university, however, noted the values at stake by pointing to the value of "academic freedom and unfettered curiosity" as "prime tenets of our quest to turn research into practical solutions for local and global needs" (Harrisburg University, 2020).

The core of this issue is that a paper slated for presentation at a conference was suppressed before its scientific credentials could be properly evaluated. The argument used to this end was that crime prediction technologies "reproduce injustices and causes real harm" (Coalition for Critical Technology, 2020).

REFUTATION OF ARGUMENT A

The first premise involves a moral evaluation of the implications of Y. According to the framework we have established, such evaluations belong to the domain of ethics. The second premise belongs to the domain of politics, as it expresses a desire for regulating science. However, the premise assumes that ethicists should determine what kind of science should be conducted.[5] The third premise is considered a factual premise, which is either true or false and which becomes the premise by which we determine whether any scientific activity Y should be suppressed. The problematic premise in this argument is P2, and with a proper division of science, ethics, and politics, such a premise becomes untenable. Without this premise, the conclusion is not necessarily that X should be suppressed. Instead, we propose an alternative premise that recognizes the division between the domains in question.

ALTERNATIVE PREMISE 2:

• The application of X should be regulated according to the political will of a given community and a given context.

Context is everything. Sex robots may appear problematic if they reinforce existing misogynistic behaviors, for example, but acceptable if they help persons with disabilities. With this premise, the scientist should be free to conduct research, the ethicist should evaluate their work as morally problematic, and the political domain should be the final arbiter in questions concerning the applications of the results of X. Granted, the political will of any given population will inevitably reflect their moral inclinations, which may vary over time, but this becomes less problematic as the premise state that the application should be regulated, not that the pursuit of knowledge related to X should be prohibited.

B: "RESTRICT BECAUSE IT COULD BE USED UNETHICALLY"

The second reason some argue for the restriction of scientific activity is that some of its consequences might be negatively evaluated by part, or all, of society. The argument is somewhat similar to the precautionary principle in that it does not presume there will necessarily be negative consequences. The mere possibility of negative consequences suffices.

p_1 Theory/approach X has unknown consequences
p_2 Some of the potential concequences of theory X are morally problematic
p_3 Science which may lead to morally problematic applications should not be pursued
q Theory/approach X should be suppressed

THE ARGUMENT

The third premise might also be exchanged for a more specific variety that involves balancing the assumed probabilities of negative consequences with the perceived benefits. An example would be lower-limb exoskeletons that were first developed for warfare that then proved to

be useful for rehabilitation contexts (Fosch-Villaronga, 2019). Such a premise will not, however, change the core of the argument. More recently, IBM and Amazon have chosen to halt facial recognition software development for police use due to the discriminatory effects of current facial recognition technology (Amazon, 2020; IBM, 2020). Still, some may argue that this response is insufficient and that the technologies themselves should never have been developed. One reason for this is that we can assume there will always be someone willing to deploy technologies with profitable potential if there are legal avenues to pursue. Hence, another example concerns the facial recognition company Clearview, which sells facial recognition services to law enforcement. They recently did the opposite of Amazon and IBM, canceling all non-law enforcement contracts (Mac, Haskins, & McDonald, 2020).

Some view technology as neutral and argue that only its application can be morally evaluated. Others, however, might point out how specific technologies are inherently problematic and even oppressive, and that technologies such as machine learning and facial recognition will inevitably perpetuate existing injustice and the foundations of structural racism (Bacchini & Lorusso, 2019; Buolamwini & Gebru, 2018; Noble, 2018). It is considered impossible to prevent these technologies from being used as technologies of control, and they thus conclude that technologies with a profoundly problematic potential should not be pursued. However, potential beneficial applications of these technologies could help blind users communicate more effectively. Context should play a role in the "purpose limitation" of these technologies, and these could be informed by ethics and enforced by politics (Fosch-Villaronga, 2019).

REFUTATION OF B

The first premise is factual and uncontroversial. The second premise involves the judgment of the ethicist concerning a set of possible consequences resulting from X. As was the case with the second premise in argument A, the third premise states that ethicists should have control over the political question of whether to allow the pursuit of X. To avoid the conflation of domains, this premise must be replaced.

ALTERNATIVE PREMISE 3:

• The application of X should be regulated according to the political will of a given community after a consideration of the possible positive and negative consequences of X

Once more, the scientist is free to pursue X, but the political domain is free to regulate the application of X. With such a premise, we trust that the ethical and political domain will be effective enough to prevent undesirable usage of such technologies. If this occurs, facial recognition technology can be pursued scientifically, and the positive applications of such technologies are allowed.

C: "RESTRICT BECAUSE IT DISTRACTS FROM SOMETHING MORE IMPORTANT"

The third argument is based on the idea that a hierarchy of challenges, or ethical bads, exists. We label this premise the great chain of ethics, based on a hierarchy similar to the great chain of being (Lovejoy, 2011). At any point in time, there are a variety of ethical challenges available for scientific attention: climate change, biodiversity loss, human rights, gender issues, structural racism, violence against women, etc. These are all issues worthy of our attention. Some are long term issues, whereas others require more immediate attention. Some authors argue that ranking these issues is possible, and the approach discussed in the example below emphasizes human rights, which are clearly anthropocentric and focused on near-term issues (Nolt, 2014). For example, according to an imagined version of a great chain of ethics, structural racism in today's society ranks higher than protecting the rights of animals, which in turn ranks higher than fighting climate change and its consequences further down the road. According to this argument, it might be wrong to devote researcher attention to problems in the lower levels of the hierarchy, because a) attention is limited, and/or b) researchers should prioritize the more important issues.

p_1 It is impossible to rank ethical bads
p_2 Issue Z is ranked higher in terms of importance than issue Y
p_3 Theory/approach X focuses on issue Y
p_4 Scientific activity is a scarce good
p_5 Public attention is a scarce good
p_5 Science which detracts from combatting the worst ethical bad should not be conducted
q Theory/approach X should be suppressed

THE ARGUMENT

While such an argument might seem far-fetched, Birhane and Van Dijk (2020a, 2020b) argue that discussions of robot rights "diverts moral philosophy away from the pressing matter of the oppressive use of AI technology against vulnerable groups in society." While the authors attempt to refute those who argue that robot rights are at least conceivable, they go beyond trying to settle an academic disagreement when they call robot rights "perverse" and dismiss this philosophical discussion as "essentially science fiction." They then proceed to state that there are "altogether different ethical concerns that have to do with the democratic distribution of power" in AI and that human beings are the "real challenge for AI ethics" (Birhane & Van Dijk, 2020a).

REFUTATION OF C

This argument involves not just conflation between the domain of science, ethics, and politics but also a fundamentally objectionable perception of the domain of ethics itself. The first and second premises are based on the assumption that it is possible to discover, for example, that fighting structural racism should be prioritized over mitigating climate change. Such a view of ethics can never be the basis of scientific policy because it is as ill-equipped to unite people in an agreement on the basis of moral evaluations as is religion. As soon as people disagree on which ethical concerns are most pressing, the argument crumbles unless someone desires to use the political domain to enforce a particular kind of ethical view. This would, of course, be an apparent conflation of ethics and politics.

The third premise is a factual statement about X and is uncontroversial. Premises 4 and 5 can be considered factually correct. It is only by the introduction of premise 6, however, that they suddenly conclude that anything other than devoting one's life and career to fighting for a particular ultimate ethical bad is legitimate. Premise 6 conflates the ethical and the political. In a pluralistic world, premises 1 and 6 are false, and premise 2 must either be taken as an axiom or be considered invalid due to premise 1 being false. This argument is faulty to such a degree that replacing a premise will not suffice – the entire argument must be replaced by an alternative that respects the three domains.

ALTERNATIVE ARGUMENT:

p_1 It is impossible to rank ethical bads
p_2 X, Y, Z constitute different threats to society and/or citizens
p_3 Politics concerns the preservation of society and/or citizens
q Politics should encourage increased understanding of X, Y, Z

In this argument, ethicists are free to condemn whatever activity they desire. When they do, and present valid arguments in favor of their claims, politics should encourage increased understanding of the activity. With a sufficient understanding in place, politics should enforce society's political will and regulate the application of the knowledge and innovation ensuing from the pursuit of increased understanding. Once more, scientists are free, even if politics encourage increased activity in specific fields.

D: "RESTRICT BECAUSE PERSON X HAS A PARTICULAR CHARACTERISTIC"

The final group of arguments, based on the idea that scientific activity should be evaluated based on who performs it, will be briefly established. There are two distinct kinds of arguments in this group:

ARGUMENT D1

p_1 Person X performs scientific activity Y
p_2 Person X is of type A
p_3 People of type B have an epimistic advantage related to Y
p_4 Science should only be conducted by those with an epistemic advantage
q Scientific activity Y should be suppressed

ARGUMENT D2

p_1 Person X performs scientific activity Y
p_2 Person X has acted unethically in a field unrelated to Y
p_3 Science should be conducted by people with good ethics
q Scientific activity Y should be suppressed

The first argument often involves the claim that certain persons or groups are epistemically privileged. For example, Nash (2008) states that "marginalized subjects have an epistemic advantage," which is a form of standpoint theory/epistemology (Anderson, 2020; Godfrey- Smith, 2003). People are situated knowers, which implies that our situations affect what we can know (Haraway, 1988; Smith, 1987). While these claims often stem from feminist literature, there are many possible categories of marginalization, and intersectionality and the compound effects of, for example, gender and race, is thus of importance (Collins, 2002). While Nash does not state that those with this knowledge are the only ones that should be allowed to speak on a subject, the implications of epistemic privilege could suggest that those without said privilege can be discounted and dismissed. The second argument states that the ethical conduct of scientists should determine whether their science should be suppressed. On this account, a person's scientific merit will carry little weight if the person has acted unethically, even if this is entirely unrelated to his scientific activity.

The first argument might involve research on structural racism in algorithmic governance. Let us assume that the third premise is based on standpoint theory and that it involves the claim that only people who are subject to racism are epistemologically privileged. Person X is light-skinned and thus not privileged. Their research should be suppressed due to their inability to understand the topic. Furthermore, some might add, X is privileged and will inevitably reinforce and re-establish the power structures at the core of the problem.

The second argument involves scientist Y. Scientist Y just achieved a break-through in facial recognition and neural interface technology, al owing people with prosopagnosia[6] to see with a new kind of special glasses. At the same time, he made a public statement about women and dark-skinned people that is considered unethical and morally problematic according to the people in charge of determining the truth content of premise two in D2. According to this argument, Y's science should be suppressed.

REFUTATION OF D

The first kind of argument in group D is related to specific groups/people's epistemic advantage. Some might argue that P3 is problematic and that universalism implies that we are epistemically equal. Epistemic privilege is ubiquitous, and different people's backgrounds and positions let them see certain aspects of a phenomenon differently from others.

Moving from P3 to P4 involves a grave danger, however. While some people may be epistemically privileged, there is no way to determine on an individual basis how this affects persons conducting a scientific activity. A person from a group that is considered unprivileged may easily and more profoundly provide new insight into a phenomenon that may escape the privileged. However, and if we believe in universalism, this is not reason enough to restrict individuals or groups in science based on what type they are or their characteristics. Furthermore, ranking the various epistemic advantages and agreeing on which ones should matter seem close to impossible. The norm of universalism clearly shows how group D arguments, which include characteristics of the scientist in the evaluation of science, conflict with these norms.

As a result, premise 4 should be stricken, and once it is, the conclusion falls. We might encourage the epistemically privileged to research Y, but we must also be open to the unprivileged voices.

p_1 Person X performs scientific activity Y
p_2 Person X has acted unethically in a field unrelated to Y
p_3 Science should be conducted by people with good ethics
q Scientific activity Y should be suppressed

The second kind of argument in group D was the following: If we accept universalism this argument is perhaps the easiest to refute, as P3 violates the division between knowledge and values – science and ethics. If we value science as a systematic quest for new knowledge, it makes little sense to couple scientific insight with the morality of whoever derived this insight. Doing so would not only lead to intentional blindness to potential truths and the positive potential of science, but it would also make any form of science impossible if we include the history of science, including historical injustice, outright discrimination and unequal access to it.

While Y's scientific activity should not be suppressed, we stress that scientists do not have a right to commit crimes or break an employer's ethical guidelines with impudence. Furthermore, they have no claim to public admiration, should they choose to conduct themselves in ways that society disapproves of. Still, their scientific activity, and their findings, should not be evaluated on the ethical conduct concerning who they are.

DISCUSSION

The ideas behind all four original arguments were based on an inherent conflation of the domains of science, ethics, and politics. For various reasons, we have rejected the original form of these arguments and instead proposed varieties that respect the demarcation of science, ethics, and politics to ensure a better assessment of the potentially adverse consequences of research on AI.

This is similar to the argument proposed by Carr (2011), that scientists are often not in the best position to evaluate the ethical consequences of their work, and that their expertise is in their disciplines, not in ethics. For example, the contemporary understanding of Responsible Research and Innovation (RRI) promotes reflection upon the consequences and outcomes of technological research and development (R&D) that foster the incorporation of societally-oriented considerations into the research or the design process (Stahl, McBride, Wakunuma, & Flick, 2014). These exercises enable researchers and designers working in this area to: (1) anticipate the potentially adverse consequences of their work to build socially robust and risk-free research; (2) reflect mindfully about their work, framing issues, problems, and proposed solutions; (3) be inclusive and conduct research not only for society but also with society, thus involving a wide range of stakeholders from the early stages of the process; (4) respond to circumstances that no longer align with society's continually evolving needs and public values; and (5) be transparent about the research to enable public scrutiny and dialogue.

However, while these strategies may help developers see what these public values are, what purpose these technologies serve to society, and how existing relationships will change, it may not be desirable to create a system in which scientists are left to pursue these efforts alone. Ideally, scientists should be free to pursue science following a precautionary approach that respects fundamental rights. Instead of becoming a jack of all trades (and master of none) that concentrates power and knowledge about all disciplines, their efforts should open avenues for collaboration that allow for a conflation of disciplines, blurring the capacity to critically assess the boundaries, limits, and opportunities of such advancements. Therefore, ethicists and politicians must continuously and

actively evaluate and regulate the science produced so that it is not uncritically applied in society.

Sattarov (2019) argues that power is so intertwined with science that it necessitates more political control in science, and that scientists should also become ethicists. We, on the other hand, argue that this requires that the domain of politics rises to the challenge and sufficiently regulates the domain of science and business. Much research on AI happens in private corporations, and we must not take it for granted that Big Tech should be allowed to conduct their research and experiments on personality profiles, facial recognition, and nudging, for example, in the wild, as they tend to do today (Zuboff, 2019). While Cohen (2019) argues that Big Tech is not some unregulated wild west, it seems clear that stricter and more proactive regulation is certainly possible. In light of the potential adverse effects of AI highlighted in this article, such an approach may also be desirable.

Furthermore, politician will not see themselves as purveyors of ethical truths, but, rather, as one whose duty it is to uphold order and allow a pluralism of ethical beliefs and scientific activity to thrive. However, one major problem remains if one adheres to such a view of science: how do we control its adverse effects? As noted: If science is to be free, and if we do not demand that the scientists or, say, Big Tech companies, are themselves ethical, this necessitates a robust and more active government that regulates and makes policies based on political processes and the work of ethicists that analyze the implications of scientific progress. For a society to flourish, science must be free. However, this does not mean that science's application should be free, as it falls under the domains of ethics and politics and the application of liberal principles. Nor does it mean that science should be completely unrestricted, as, for example, research on dangerous viruses, etc., must clearly be regulated and under societal control. Our point is that if society deems, for example, facial recognition, to be dangerous in a particular context, it has the right to restrict the application of said technologies through politics and regulation (Sætra, 2021a). However, it should not prevent the domain of science from producing as much knowledge and insight into the phenomena as possible at the risk of foregoing the benefits it may entail, and even the means to prevent other negative outcomes. That is the only way towards a proper evaluation of whether we should apply it, and, not least, realize the positive potential of such technologies.

Of great importance is how science's norms and theories also create a clear imperative for positive and forceful political activity, particularly concerning the norm of universalism and the marketplace of ideas. If a specific group in society is disadvantaged and not well represented in science, such as the female population, the LGBT community, or persons with disabilities, the market of ideas will not function effectively (Hadorn, 1992; Gibney, 2019; Nature, 2020). In this sense, the abundance of theoretical equality of opportunity is not sufficient, as historical injustice and structural racism may recreate barriers to equal participation in practice (Dirth, & ranscombe, 2017; Huebner, Kras, & Pleggenkuhle, 2019). However, while this is a fundamental problem for science, it is not a problem to be solved by the domain of science. A problem highlighted by ethical appraisal should be accompanied by a solid governance response, as in the EU's case with the importance of gender and sex in research and direct intervention in science. Furthermore, if we respect the division of the three domains, such a political solution should be geared towards using social reform to create a society in which structural conditions do not exclude particular groups from science on a systematic basis.

However, the dangers of a free market of anything - ideas included - are not the only danger to be aware of. Science is not readily available to the public, and totalitarian forces can easily be both more comprehensible and attractive to a populace than the diverse and often abstract ideals of liberalism and free science (Merton, 1942). This is one clear reason to be wary of the political activism and conflation partly suggested by Sattarov (2019). Strauss (1988) similarly *notes that persecution and oppression follows when a „compulsion to coordinate speech with such views as the government believes to be expedient" reigns, and this can also be related to the danger of a tyranny of the majority – a phenomenon relevant both in the age of Tocqueville (2004) and in the age of AI (Sætra, 2019a). Strong institutions, democracy, and vibrant and free domains of both ethics and science may support our liberal democracies' scaffolding. If we believe that democracy and liberalism are the solution and way forward, science helps demonstrate this, and ethics will show and explain why. By restricting any of these domains, we begin dismantling the very ideals of liberty and toleration that we may have aimed to protect by restricting science.

The separation of science from ethics and politics requires great diligence by the ethicist and politician. By upholding the distinction, a balance of power is created, and each domain's roles become crucial for preventing adverse outcomes. One possible negative outcome is that we lose oversight and control over the application of science, and we argue that this has partly occurred as regulators have allowed the growth of research on AI in the private sector and Big Tech, which is largely unregulated compared to academia. Part of the danger stems from faulty science, and it is essential for the domains of ethics and politics to continuously work on exposing such faults. This is partly related to enforcing strict research ethics and ensuring the norms of openness and communism, enabling transparency and the possibility of monitoring science (Collingridge, 1980; Merton, 1942). Such ethics must not only be applied to public research and research in

academia; it must also be implemented in the domain of science more generally – public and private.

Many of the examples listed above involve a desire to restrict scientific activity in order to achieve justice. As noted by Thrasher (2012), justice „is about conflict" – resolving and reducing it – but conflict can be reduced in many ways, and one way to achieve peace and solve a conflict is to „kill off all those who disagree." This approach he labels justice as a victory, and it is a Thrasymacian view of justice (Thrasher, 2012). This is akin to silencing the voices we disagree with and using either individual or collective power to police and restrict the world of science while imposing subjective labels of what is, and what is not, legitimate science and regulating what are considered legitimate questions for scientists. Justice and conflict could also be resolved by respecting the liberty of scientists and the pluralism of values. It is based on debate and the idea that reasonable disagreement is ubiquitous. Such disagreement must be dealt with by debate and the construction of arguments in favor of one's own beliefs or attempts to refute others' perceived beliefs (Thrasher, 2012). This is the proper way of science, and it is a world apart from the desire to silence and eliminate uncomfortable ideas.

1 See http://inbots.eu/.
2 See http://www.unesco.org/new/en/natural-sciences/science-technology/engineering/engineering-education/.
3 See https://www.stopkillerrobots.org/.
4 For the current undertaking, we use Merton's (1942) norms as shorthand for the more comprehensive ethos. These are the four "institutional imperatives": Universalism, communism, disinterested and organized skepticism. These norms are promulgated through „prescriptions, proscriptions, preferences, and permissions" (Merton, 1942). The norms have subsequently been developed, adjusted and discussed, but it is Merton's original contribution that is here considered. See Ziman (2002) and Macfarlane and Cheng (2008) for more details on what is often referred to as Mertonian norms.
5 When politics delegate the authority to evaluate the safety of a project to ethicists, this involves ethicists acting on a mandate from politics, and this is often both necessary and unproblematic.
6 Describes an inability to discriminate between faces – also referred to as face blindness.

CONCLUSION:

AI has already become the next big thing. With its incredible scientific advances, various forms of regulations and ethical guidelines proliferate, leading to a questionable and unclear conflation of science, ethics, and politics. While mechanisms to anticipate, control, and steer AI development to prevent adverse consequences seem necessary, the lack of a clear balance of power and clearly defined roles between science, ethics, and politics is causing much confusion.

In this article, we have examined a set of arguments aimed at restricting scientific activity. These are based on the idea that ethical and political considerations must, to some degree, restrict science. These arguments are based on a potentially dangerous conflation of science, ethics, and politics currently present in AI development and discourse, which blurs and distorts the liberty and responsibility scientists should have to pursue their research. In this respect, the role of ethics in relation to science and law remain very much unclear (González-Fuster, 2020).

While moral philosophers should criticize, uncover, and highlight all kinds of problems related to scientific progress, a conflation of science, ethics, and politics can provoke undesirable outcomes. While there are apparent reasons for politics to be concerned with injustice in science, based on, for example, unequal representation, suppressing science and directly controlling it may be premature and misguided. Instead, a better course of action would be to work towards building a just society with equal opportunities where the different domains serve their original control functions to adequately frame and guide progress. These insights stem from traditional theories from the philosophy and sociology of science and from liberal political theory.

The article aimed to examine whether AI has changed the situation in a way that warrants a conflation between science, ethics, and politics. However, while we have examined several arguments in favor of such a position, we have concluded that such arguments may be erroneous. Given the potential of AI, its development and usage may be best handled by upholding the different functions of the three domains to ensure that it truly benefits society.

REFERENCES:

Amazon. (2020). We are implementing a one-year moratorium on police use of Rekognition. Retrieved from https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition

Anderson, E. (2020). Feminist epistemology and philosophy of science. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2020 ed.).

Bacchini, F., & Lorusso, L. (2019). Race, again: how face recognition technology reinforces racial discrimination. Journal of Information, Communication and Ethics in Society.

Bauman, Z. (2013). Liquid love: On the frailty of human bonds. John Wiley & Sons.

BBC News. (2019, October 19th). Barack Obama challenges ‚woke' culture. BBC News. Retrieved from https://www.bbc.com/news/world-us-canada-50239261
Birhane, A., & Van Dijk, J. (2020a). A Misdirected Application Of AI Ethics. Noema. Retrieved from https://www.noemamag.com/a-misdirected-application-of-ai-ethics/

Birhane, A., & van Dijk, J. (2020). Robot Rights? Let's Talk about Human Welfare Instead. Paper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.

Boocher, S. (2012). Environmental justice. In D. Schmidtz (Ed.), Technology and What do Do About It. New York: Oxford University Press.

Brundage, M., & Bryson, J. (2016). Smart policies for artificial intelligence. arXiv preprint arXiv:1608.08196
Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: the legal lacuna of synthetic persons. Artificial Intelligence and Law, 25(3), 273-291.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Paper presented at the Conference on fairness, accountability and transparency.

ChoGlueck, C. (2018). The error is in the gap: Synthesizing accounts for societal values in science. Philosophy of science, 85(4), 704-725.

Coalition for Critical Technology. (2020). Abolish the #TechToPrisonPipeline. Retrieved from https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. Ethics and Information Technology, 12(3), 209-221.

Cohen, J. E. (2019). Review of Zuboff's The Age of Surveillance Capitalism. Surveillance & Society, 17(1/2), 240-245.

Collingridge, D. (1980). The social control of technology. London: Frances Pinter.

Collins, P. H. (2002). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge.

Danaher, J. (2019). Regulating Child Sex Robots: Restriction or Experimentation? Medical Law Review, 27(4), 553-575.

Dirth, T. P., & Branscombe, N. R. (2017). Disability models affect disability policy support through awareness of structural discrimination. Journal of Social Issues, 73(2), 413-442.

Feyerabend, P. (1970). Consolations for the Specialist. Criticism and the Growth of Knowledge, 4, 197-230.

Feyerabend, P. (1993). Against method. London: Verso.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.

Foer, F. (2017). World without mind: Random House.

Fosch-Villaronga, E. (2019). Robots, healthcare, and the law: Regulating automation in personal care. Routledge.

Fosch-Villaronga, E., & Heldeweg, M. (2018). " Regulation, I presume?" said the robot–Towards an iterative regulatory process for robot governance. Computer law & security review, 34(6), 1258-1277.

Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2020) Don't guess my gender, gurl: The inadvertent impact of gender inferences. BIAS 2020: Bias and Fairness in AI Workshop at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 14-18 September 2020, online.

Fosch-Villaronga, E., & Poulsen, A. (2020). Sex care robots: Exploring the potential use of sexual robot technologies for disabled and elder care. Paladyn, Journal of Behavioral Robotics, 11(1), 1-18.
Gellers, J. (2020). Rights for Robots: Artificial Intelligence, Animal and Environmental Law. bingdon: Routledge.

Gibney, E. (2019). Discrimination drives LGBT+ scientists to think about quitting. Nature, 571(7763), 16-18.

Gomes, A., Antonialli, D. & Dias-Oliva, T. (2019). Drag queens and Artificial Intelligence: should computers decide what is 'toxic' on the internet? Internet Lab. Retrieved from http://www.internetlab.org.br/en/freedom-of-expression/drag-queens-and-artificial-intelligence-should-computers-decide-what-is-toxic-on-the-internet/ (last accessed 30 December 2020).

Grush, L. (2015). Google engineer apologizes after Photos app tags two black people as gorillas. The Verge. Retrieved from https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas (last accessed 30 December 2020).

Gunkel, D. J. (2017). Mind the gap: responsible robotics and the problem of responsibility. Ethics and Information Technology, 1-14.

Gunkel, D. J. (2018). Robot rights. London: MIT Press.
Hadorn, D. C. (1992). The problem of discrimination in health care priority setting. JAMA, 268(11), 1454-1459.

Hao, K. (2020, June 23rd). AI researchers say scientific publishers help perpetuate racist algorithms. MIT Technology Review. Retrieved from https://www.technologyreview.com/2020/06/23/1004333/ai-science-publishers-perpetuate-racist-face-recognition/

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist studies, 14(3), 575-599.

Harrisburg University. (2020, May 5th). Facial recognition software paper not being published. Retrieved from https://harrisburgu.edu/hu-facial recognition-software-identifies-potential-criminals/

Huebner, B. M., Kras, K. R., & Pleggenkuhle, B. (2019). Structural discrimination and social stigma among individuals incarcerated for sexual offenses: Reentry across the rural–urban continuum. Criminology, 57(4), 715-738.

IBM. (2020). IBM CEO's Letter to Congress on Racial Justice Reform. Retrieved from https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/.

Johnston, S. F. (2018). The technological fix as social cure-all: origins and implications. IEEE Technology and Society Magazine, 37(1), 47-54.

Kitcher, P. (2003). Science, truth, and democracy: Oxford University Press.

Kourany, J. A. (2003). A philosophy of science for the twenty-first century. Philosophy of science, 70(1), 1-14.
Latour, B. (1999). Pandora's hope: essays on the reality of science studies. Cambridge: Harvard university press.

Levy, D. (2009). Love and sex with robots: The evolution of human-robot relationships: New York.

Longino, H. E. (1990). Science as social knowledge: Values and objectivity in scientific inquiry: Princeton University Press.

Lovejoy, A. O. (2011). The great chain of being: A study of the history of an idea: Transaction Publishers.
Mac, R., Haskins, C., & McDonald, L. (2020, May 7th). Clearview AI Has Promised To Cancel All Relationships With Private Companies. Buzzfeed News. Retrieved from https://www.buzzfeednews.com/article/ryanmac/clearview-ai-no-facial-recognition-private-companies

Macfarlane, B., & Cheng, M. (2008). Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms. Journal of Academic Ethics, 6(1), 67-78.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology, 6(3), 175-183.

Merton, R. K. (1942). Science and technology in a* democratic order. Journal of legal and political sociology, 1(1), 115-126.

Merton, R. K. (1973). The normative structure of science. . In N. W. Storer (Ed.), The sociology of science: Theoretical and Empirical Investigations. Chicago: The University of Chicago Press.

Meynaud, J. (1969). Technocracy. New York: Free Press.
Mill, J. S. (1985). On Liberty. London: Penguin books.

Mulkay, M. J. (1976). Norms and ideology in science. Social Science Information, 15(4-5), 637-656.

Nash, J. C. (2008). Re-thinking intersectionality. Feminist review, 89(1), 1-15.

Nature (2020) Accounting for sex and gender makes for better science. Editorial, Nature 588, 196, https://doi.org/10.1038/d41586-020-03459-y.

Nisevic, M., Fosch-Villaronga, E., Sears, A. M., Custers, B. H. M. (2021) Automated Decision-Making under EU Data Protection Law. In: Kosta, E., & Leenes, R. (2021) Research handbook on EU data protection. Edward Elgar Publishing, forthcoming.

Nolt, J. (2014). Environmental ethics for the long term: An introduction: Routledge.

Noorman, M., Keymolen, E., Schellekens, M. de Groot, A., Conca, S., Coenmans, E., Leenes, R., Zhao, B. (...) Taylor, L. (2019) Response on the draft ethical guidelines for trustworthy AI produced by the European Commission's High-Level Expert Group on Artificial Intelligence. The AI and Robotics group at the Tilburg Institute for Law, Technology and Society. Retrieved from: https://www.tilburguniversity.edu/sites/default/files/download/NotesonHLEGDraftethicalguidelines_30012019_1.pdf

Næss, A. (1989). Ecology, community and lifestyle: outline of an ecosophy: Cambridge university press.
Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

Poulsen, A., Fosch-Villaronga, E., & Søraa, R. A. (2020). Queering machines. Nature Machine Intelligence, 2(3), 152-152.

Redding, R. E. (2013). Politicized science. Society, 50(5), 439-446.

Smith, D. E. (1987). The everyday world as problematic: A feminist sociology: University of Toronto Press.

Strauss, L. (1988). Persecution and the Art of Writing. Chicago: University of Chicago Press.

Sætra, H. S. (2019a). The tyranny of perceived opinion: Freedom and information in the era of big data. Technology in Society, 59, 101155.

Sætra, H. S. (2019b). When nudge comes to shove: Liberty and nudging in the era of big data. Technology in Society, 59, 101130.

Sætra, H. S. (2020a). First, They Came for the Old and Demented. Human Arenas, 1-19. doi:https://doi.org/10.1007/s42087-020-00125-7

Sætra, H. S. (2020b). The Parasitic Nature of Social AI: Sharing Minds with the Mindless. Integrative Psychological and Behavioral Science, 1-19.

Sætra, H. S. (2020c). A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government. Technology in Society, 101283.

Sætra, H. S. (2021a). Confounding Complexity of Machine Action: A Hobbesian Account of Machine Responsibility. International Journal of Technoethics, 12(1).

Sætra, H. S. (2021b). Loving robots changing love: Towards a practical deficiency-love. Journal of future robot life.

Sætra, H. S. (Forthcoming). The AI ethicist's dilemma: Fighting Big Tech by supporting Big Tech. TBA.

Sattarov, F. (2019). Power and Technology: A Philosophical and Ethical Analysis: Rowman & Littlefield.

Sparrow, R. (2007). Killer robots. Journal of applied philosophy, 24(1), 62-77.

Stahl, B. C., McBride, N., Wakunuma, K., & Flick, C. (2014). The empathic care robot: A prototype of responsible research and innovation. Technological Forecasting and Social Change, 84, 74-85.

Sullins, J. P. (2012). Robots, love, and sex: the ethics of building a love machine. IEEE Transactions on Affective Computing, 3(4), 398-409.

Tocqueville, A. D. (2004). Democracy in america. New York: The Library of America.

Thrasher, J. (2012). Environmental justice. In D. Schmidtz (Ed.), Environmental ethics: what really matters, what really works. New York: Oxford University Press.

Vincent, M. (2018) Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech. The Verge. Retrieved from https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai (last accessed 30 December 2020).

Walker, K., & Wan, F. (2012). The harm of symbolic actions and green-washing: Corporate actions and communications on environmental performance and their financial implications. Journal of business ethics, 109(2), 227-242.

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. Information, Communication & Society, 20(1), 118-136.

Ziman, J. (2002). Real science: What it is and what it means: Cambridge University Press.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019. New York: PublicAffairs.