

Ethical Foresight Analysis: What it is and Why it is Needed?

Luciano Floridi^{1,2} Andrew Strait¹

Abstract

An increasing number of technology firms are implementing processes to identify and evaluate the ethical risks of their systems and products. A key part of these review processes is to foresee potential impacts of these technologies on different groups of users. In this article, we use the expression Ethical Foresight Analysis (EFA) to refer to a variety of analytical strategies for anticipating or predicting the ethical issues that new technological artefacts, services, and applications may raise. This article examines several existing EFA methodologies currently in use. It identifies the purposes of ethical foresight, the kinds of methods that current methodologies employ, and the strengths and weaknesses of each of these current approaches. The conclusion is that a new kind of foresight analysis on the ethics of emerging technologies is both feasible and urgently needed.

Keywords Artificial intelligence · Data science · Ethics · Foresight analysis · Machine learning · Innovation

1 Introduction

In this article, we use the expression Ethical Foresight Analysis (EFA) to refer to a variety of analytical methodologies for anticipating or predicting the ethical issues that technological artefacts may raise. In the following pages, we summarise the purpose, strengths, and weaknesses of six commonly-used forms of foresight analysis that have been applied in evaluations of technological artefacts. The article is structured in six sections. In Sect. 2, we provide a brief history of foresight analysis and an evaluation of its major concepts to situate EFA within the relevant sociological, philosophical, and scientific fields of study. In Sect. 3, we review six EFA methodologies and several related sub-methodologies, commenting on the intended purposes, uses, strengths, and weaknesses of each. In Sect 4, we discuss the known

✉ Luciano Floridi
luciano.floridi@oii.ox.ac.uk

¹ Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

² The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK

limitations of current EFA methodologies as applied within technology companies. In Sect. 5, we suggest potential future approaches to EFA that are worthy of investigation. In the conclusion, we recommend the development of a more focused EFA methodology for technology firms to employ in an ethical review process.

2 Background

The literature on ethical foresight analysis often uses different terms and frameworks to describe similar concepts. In this first section, we provide some common background in order to avoid potential misunderstandings.

2.1 Definitions

In this paper, we adopt the definition of a “technology” as “a collection of techniques that are related to each other because of a common purpose, domain, or formal or functional features” (Brey 2012). An “artefact”, by contrast, refers to a physical or digital product, service, or platform created out of a technological field that produces a desired result. An “emerging” technological artefact refers to one that is in its design, research, development, or experimental stages, including beta testing.

2.2 A Brief History of Foresight Analysis

Approaches to Ethical Foresight Analysis stem from the broader field of Foresight Analysis (FA), which has been used since the 1950s for anticipating or predicting the outcome of potential policy decisions, emerging technologies and artefacts, or economic/societal trends (Bell 2017). At its outset, FA was commonly used for the purpose of government planning, business strategy, and industry development, including environmental impact analyses and predictions of economic growth. Delphi (more on this presently) was one common FA methodology used for government and economic planning (Turoff and Linstone 1975). These methodologies are characterized by their use of a variety of quantitative and qualitative methods to predict potential future scenarios, including surveys, interviews, focus groups, scenario-analysis, and statistical modelling.

EFA also draws upon the field of Future-Oriented Technology Analysis (FOTA), which itself stems from the fields of innovation studies in the 1970s and 1980s. These fields sought to apply forecasting models to the specific task of identifying the characteristics and effects of future technological advancements (Miles 2010). FOTA methodologies draw heavily on concepts from the field of Science and Technology Studies (STS), which focuses on the interplay between technologies and different actors in a society (Pinch and Bijker 1984). There are several, similarly-named, sub-categories of FOTA, which seek to accomplish different ends (Nazarko 2017). We briefly discuss here four of them. *Future Studies* focuses on forecasting what possible or probable technologies may exist in the future and how those technologies may be used by a society. *Technology Forecasting* attempts to predict

future *characteristics* of a particular technology. *Technology Assessment* evaluates the impacts of an existing or emerging technological artefact on an industry, environment, or society, and can include recommendations for actions to mitigate risks or improve outcomes. Lastly, *Technology Foresight* seeks to build a long-term vision of an entire technology, economic, or scientific sector by identifying strategic areas of research to improve overall welfare (Bell 2017; Tran and Daim 2008).

When compared to Future Studies and FOTA, Ethical Foresight Analysis (EFA) has a slightly different purpose: forecasting “ethical issues at the research, development, and introduction stages of technology development through [the] anticipation of possible future devices, applications, and social consequences” (Brey 2012). Thus, EFA seeks to identify what kinds of ethical issues an emerging technology or artefact will raise in the future, with some methodologies focused on long term forecasting and others focusing on shorter term projections.

2.3 Relevant Concepts for Ethical Foresight Analysis

EFA is rooted in several foundational concepts from the field of Science and Technology Studies (STS). One important idea that underlies EFA is that the uses and effects of a technological artefact on society are not determined solely by the artefact’s design but also by different “relevant social groups” who co-opt that technology for specific societal needs. These needs may be different from the intended use of the developer (Collins 1981; Mulkay 1979). For example, the initial main application of the telephone was thought to be that of broadcasting music. As Fischer (1992), 5 puts it in his analysis of the telephone’s development in the US, while a new technology may alter “the conditions of daily life, it does not determine the basic character of that life—instead, people turn new devices to various purposes.” Technological artefacts and society thus “co-evolve” to establish the ultimate uses of an artefact (Lucivero 2016). To take another example, the 19th century bicycle began as a tall front wheeled device with rubber tires but developed into customized models that fit the needs of different relevant social groups. Younger riders, for example, wanted a bicycle with a more forward-facing frame they could use to race competitively while the general public desired a safer design that included a cushioned seat. These desired uses for the bicycle resulted in the development of different types of frames for each relevant social group and the common adoption of an inflatable tire to maximize speed and create more shock absorption (Pinch and Bijker 1984).

Other key concepts from STS relevant to EFA are those of *disruption*, *stabilization*, and *closure*. Disruption refers to the period of time when a technology or artefact is introduced, characterized by rapid changes in both the technology’s development and uncertainty from society about how a technology or artefact should be used. This is the period of time when interpretive flexibility, or the capability of relevant social groups to impart different meanings, expectations, and uses of a technological artefact, is at its highest. Stabilization and closure refer to the periods in which an artefact’s use amongst members of a society settles and eventually becomes well accepted amongst developers and users (Pinch and Bijker 1984).

With emerging technologies, the initial period of disruption produces the greatest obstacle that ethical foresight methodologies attempt to solve: *uncertainty* (Cantin and Michel 2003). If a developer cannot envision the actual uses of their technology, it can be difficult fully to understand its ethical impact. For that reason, uncertainty is especially high in the R&D stages, when the specific functions and features of a technology are still unclear. Uncertainty also characterises how a technology may affect and change existing moral standards that would then affect its impact. Moral concepts such as autonomy and privacy may, for example, shift over time as technologies and artefacts change how a society defines ‘private space’ and the ‘self’ (Lucivero 2016). A smartphone app that seeks to use health data for diagnosing medical conditions could encounter resistance on the grounds that it invades privacy; alternatively, it may enjoy success by shifting how a society views the sensitive nature of health data in light of the overall medical benefits the app provides. EFA methodologies seek to resolve the uncertainty about which of those outcomes is more likely.

A final foundational concept of EFA from the field of Technology Assessment refers to the difficulty of measuring *moral change*. As Boenink et al. (2010) note, moral change occurs on several different levels. “Macro” change refers to extremely slow and gradual changes in abstract moral principles within different social groups and societies, while “meso” changes refer to alternations of institutional norms and practices (such as the concept of “informed consent” in modern privacy policies). “Micro” changes refer to niche moral issues occurring in local circumstances where negotiation and change frequently occur (Boenink et al. 2010). Successful EFA methodologies tend to forecast micro changes with higher reliability than meso and macro changes, as an ethicist has less complex unknown variables to account for in an ethical analysis.

2.4 When is Ethical Foresight Analysis Useful?

Over the past several years, technology firms and research labs have increasingly sought to implement various forms of ethical review of their products, services, and research. These review processes often pursue two goals: (1) to analyse the ethical, political, or societal effects of a current or emerging technological artefact; and (2) to identify actionable ways to mitigate any risks of harm an artefact poses (e.g. through changes to product features, research design, or the cancellation of a project altogether). A common challenge in these review processes, particularly for emerging technologies, is the difficulty of holistically identifying and prioritizing potential harmful impacts of an artefact given the complex ecosystems and shifting socio-political environments into which a technology is deployed. Unlike legal reviews that evaluate an emerging technology against the language of existing legal codes, ethical evaluations require articulating what the moral harms of a particular technology are and how these may change over time.

EFA approaches seek to provide grounded methods for technology and research organisations to identify the ethical implications of an emerging artefact or technology, thereby informing recommendations for mitigating those harms. Importantly,

these approaches should be understood as one tool in the toolbox for an ethical review process, not as a holistic or exhaustive form of ethical review in itself. When combined with other methods for ethical design, EFA can be a complementary structured method for assessing the risks of a particular technology.

3 Existing Methodologies of Ethical Foresight Analysis

We now turn to six current methodologies of Foresight Analysis that have been used to identify potential ethical problems with emerging technologies. The methods discussed below were identified as part of a literature review conducted in early 2018 into the most commonly cited methods of EFA. One method discussed below (DbD) was flagged to us by a reviewer of this paper. While this list is non-exhaustive, we believe the methods chosen below accurately represent the major differences in approach and practice.

3.1 Crowdsourced Single Predictions Frameworks (Delphi and Prediction Markets)

Delphi is a qualitative form of consensus-based impact analysis that has been used in business strategy and government planning for over 70 years (Turoff and Linstone 1975). More recently, Delphi has been used to forecast potential ethical implications in areas like education, policy, and nursing (Manley 2013). The core tenets of the Delphi model involve anonymized interviews or focus groups with experts from a series of fields relating to specific topics. A Delphi analysis of government agricultural policy, for example, might start with reaching out to experts in biology, chemistry, farming, social policy, community organizations, and other organizations related to any potential relevant social group that may be impacted by the analysis. Once these experts are identified, they respond to a series of questions related to a proposed plan of action or, via more open-ended questions, to hypothesize solutions to likely issues that may arise. Feedback from individuals is anonymized and respondents are prohibited from speaking to one another to discourage cross-contamination of ideas. Once this feedback is collected, a facilitator evaluates the feedback to determine common themes and areas of agreement. Once common areas of agreement are identified, the facilitator then repeats the process of soliciting expert feedback until the list of likely outcomes and ethical issues has been narrowed down (Armstrong 2008). Various sub-methods of Delphi exist, including mini-Delphi—in which non-anonymized focus groups are used instead of interviews—and web-based Delphi, in which forum members interact anonymously in real time online (Battisti 2004). Delphi can also be combined with other methods like scenario analysis, which seeks to create likely future scenarios for a technology that provide multiple possible outcomes. Tapio (2003), for example, recommends taking common themes from respondents to envision multiple future scenarios that could occur. Delphi has been used in various forms of ethical analysis, including to characterize ethical issues raised by the use of novel biotechnologies (Millar et al. 2007). In these

instances, a variety of experts and users were consulted on intended and unintended consequences of the introduction of these technologies and their results condensed into a set of most likely scenarios.

Delphi is not the only form of crowdsourced foresight analysis. *Prediction markets* are another method that collects vast amounts of information from individuals to collate it into a single predictive data point. Prediction markets are based on the idea that a market of predicting bidders can provide accurate crowdsourced forecasts of the likelihood of certain events or outcomes. Major tech firms, like Google and Microsoft, have used prediction markets internally to predict launch dates of potential products (Chansanchai 2014; Cowgill 2005). There are currently no recorded uses of prediction markets to forecast ethical outcomes of a technology or artefact, but this practice has been used widely in the healthcare industry to predict the likelihood of potential ethical issues (Polgreen et al. 2007).

3.1.1 Evaluation

Consensus-based methods like Delphi and prediction markets carry several weaknesses when applied to ethical forecasting. First, a broad range of representatives from relevant social groups is crucial for any consensus-based forecast, which raises a significant weakness of this model for forecasts of emerging technologies. Companies are often disincentivized from engaging in this kind of analysis due to potential loss of proprietary information of an emerging technology in its research and development stage. Furthermore, Delphi and prediction markets rely on the presumption that different relevant social groups are capable of understanding what a technology or artefact actually does and therefore what ethical impacts are likely to occur. If there is any kind of information asymmetry—where one group of experts understand a technology in a fundamentally different way from another—this may spoil the effectiveness of any kind of insight from a crowdsourced analysis as the exercise will devolve into group-think.

3.2 Technology Assessment (TA)

Technology Assessment (TA) is defined as a “systematic attempt to foresee the consequences of introducing a particular [artefact] in all spheres it is likely to interact with” and the effects of that artefact on a society as it is introduced, extended, and modified (Braun 1998; Nazarko 2017). There are four general types of TA (Lucivero 2016):

- (1) *Classical*, which tends to focus on a top-down centralized approach from a group of experts to develop a final product that informs decision-makers of the potential threats of an emerging technology artefact;
- (2) *Participatory*, which involves a broad scope of relevant social groups in the process of assessing a technological artefact or product to develop consensus-based/democratic options for decision-making (van Eijndhoven 1997);

-
- (3) *Argumentative*, in which TA is an ongoing process of democratic argumentation and debate to continuously refine the development of new technologies (Van Est and Brom 2012);
 - (4) *Constructive (CTA)*, in which the goal is not to develop clear options or “end products” in TA but to create a democratic means of assisting in the development of technologies and their impact on societies. This process embraces the co-evolution of society and technology and posits that objective knowledge of how a technology will affect society cannot be accurately determined (Lucivero 2016; Jasanoff 2010; Rip et al. 1995). Rather, citizens and stakeholders assist in the development of a technology or artefact from its inception to build ethical considerations into an emerging technology (van Est and Brom 2012).

TA relies on a variety of methodological tools to conduct analyses. One common method is structural modelling of complex systems to understand the interplay between different measurable variables (Keller and Ledergerber 1998). Modelling has been used primarily for assessing the economic impact of an emerging technological artefact, and this method is not well suited to ethical forecasting due to the extreme variability and complexity of ethical issues and the near impossibility of translating ethical considerations into continuous variables.

Two other related toolsets used in TA are impact analyses, which compare likely technological developments against a checklist of known issues (Palm and Hansson 2006), and scenario analyses, which forecast likely scenarios that may develop in the future (Diffenbach 1981; Boenink et al. 2010). Other commonly used methods include risk assessment, interviews, surveys, and focus groups to identify common elements of a technology that may raise ethical concerns (Tran and Daim 2008). Constructive Technology Assessment in particular relies on a series of ongoing focus groups and interviews with relevant stakeholders to provide consistent feedback during the development of a technology.

Related to TA are other forms of impact assessments that are often used to evaluate a particular technological artefact. *Privacy Impact Assessments*, for example, provide a step-by-step approach for evaluating a particular privacy practice within a company, organization, or product (Clarke 2009). *Algorithmic Impact Assessments*, by contrast, provide a structured step-by-step checklist for evaluating the health of an algorithmic decision-making system (Reisman et al. 2018). Impact assessment checklists are more commonly used to compare a new or existing artefact against certain standards that an organisation or regulatory body has previously agreed to, but they do not necessarily need to forecast what ethical outcomes may arise in the future. As we will see below, ethical checklists also play a useful role in other ethical foresight strategies.

Another closely related method is *Real Time Technology Assessment (RTTA)*. This embraces the concept of socio-technical dialogue between producers and consumers in the process of determining what new technologies should exist. Through the use of surveys, opinion polling, and focus groups, consumers provide direct input into what kinds of technologies should be designed and into the kinds of ethical dilemmas a new technology can create. RTTA practitioners have occasionally used content analysis and survey research to track trends in perceptions of ethical

beliefs about a technology over time. Additionally, RTTA practitioners have used scenario analysis techniques to map areas of ethical concern as they develop over time (Guston and Sarewitz 2002).

A more recent and promising method of TA is Designing by Debate (DbD), which combines elements of CTA with a PAR4P (Participatory Action Research for the development of Policies), a participative method of informing stakeholders about the state of a particular technology or policy area (Ausloos et al. 2018). Developed to address several issues with contemporary R&D with few previous analogues, DbD uses a resource-intensive four-step cycle to map potential challenges with traditionally unrepresented stakeholders and reach a consensus on potential solutions:

- (1) Mapping existing normative frameworks and solutions to the research or innovation challenge;
- (2) Mapping relevant stakeholders that should participate but are often ignored, forgotten or whose interests are inadequately taken into account;
- (3) Using intensive offline sessions with online follow-up interactions, leading participative exercises to collect all stakeholders' views on research problems (e.g. through peer-to-peer, stakeholder and policy debate);
- (4) Validate and integrate the results of previous steps.

DbD also uses a set of components for situations that are 'similar to ones where full DbD cycles have already been completed and/or where such full cycles are not feasible.' (Ausloos et al. 2018). These components include creating decision trees from previous DbD cycles to show how a solution was reached, compiling a list of known stakeholders to consult regularly with on new projects, and creating practical guidelines based on previous cycles.

3.2.1 Evaluation

Technology assessment methodologies, like CTA and RTTA, have significant strengths in that they provide the clearest avenue for different voices to be heard in the development and application of emerging technological artefacts. However, there are also major weaknesses with these methods. First, there are significant disincentives for a company to use crowdsourced methods in the design of an artefact. This process takes significant amounts of resources to sustain and can radically slow down the launch process of a new emerging technology. Proprietary information becomes difficult to keep secret, which can ultimately harm a company's launch strategy. TA tends to function extremely well when used for government and industry-wide planning for the introduction and development of a technology or artefact. This is the trade-off made in most crowdsourced models: speed and secrecy of a product launch are sacrificed for greater public understanding of (as well as engagement with) the ethical issues at play. This method thus works better as a safeguard to help build products and artefacts that have ethical principles baked into them, but less well as a foresight methodology to identify what potential ethical issues may arise from a new technology. Methods like DbD provide some promising solutions

to these issues, namely offering a valuable way for organisations to streamline future ethical review cycles.

3.3 Debate-Oriented Frameworks (eTA)

Ethical Technology Assessment (eTA), developed by Elin Palm and Sven Ove Hansson, is a form of TA that seeks to forecast potential ethical issues of an emerging technology through “a continuous dialogue” between the developers of that technology. This process continues from the early stages of research development until well after the technology’s launch (Palm and Hansson 2006). The approach seeks to avoid a “crystal ball” practice of speculating about what potential outcomes may arise in the far future via one single evaluation of a technology. Rather, a sustained and continued assessment ensures that a continuous feedback of emerging ethical concerns from a technology are worked into the next iteration of the design/application process (Tran and Daim 2008).

The process of curating this dialogue involves repeated interviews or surveys of technology developers and other relevant stakeholders of an emerging technology. As opposed to a wider-scale source of stakeholders used in Delphi, stakeholders of eTA tend to be developers or other groups from the specific company or industry working on the technology in question. This dialogue is guided by a checklist of common ethical issues that experts in a particular technological area have agreed upon. Palm and Hansson list nine areas of ethical concern to frame these conversations:

- (1) Dissemination and use of information
- (2) Control, influence and power
- (3) Impact on social contact patterns
- (4) Privacy
- (5) Sustainability
- (6) Human reproduction
- (7) Gender, minorities and justice
- (8) International relations
- (9) Impact on human values.

These dialogues are meant to produce a single conception of the ethical issues that may arise in an emerging technology, which are then fed back into artefact’s R&D and design process. In evaluating the responses from these dialogues, a practitioner of eTA is encouraged not to use a single ethical framework to produce one policy recommendation. As Palm and Hansson note, different ethical frameworks (e.g. utilitarianism, deontology) can produce radically different recommendations for ethical behaviour in a particular situation. Unlike crowdsourced and consensus-based practices, the goal of eTA is to flag areas of disagreement and concern between stakeholders to create a new moral framework with which to judge the development of a technology (Grunwald 2000). The end result of this process is to

use different moral concepts to flag alternative recommendations for future iterations of the artefact or technology.

3.3.1 Evaluation

eTA has several limitations. First, like other TA methods, it is an abstract framework with little guidance as to how to source the relevant data in these dialogues or what critical thresholds are expected to be met for a successful forecast. This raises concerns for consistency, as the lack of clear guidance for a forecaster to follow may result in some eTA forecasts involving more stakeholders than others. Second, Brey (2012) notes that the checklist of nine moral problems is not exhaustive and may not reflect the ethical concerns of all moral frameworks. eTA also requires significant amounts of time and resources to analyse a technology continuously, which may raise issues in a fast-paced industry environment where artefacts undergo constant changes and tweaks.

A greater concern with eTA is the problem of an overly-narrow source for dialogue about the potential ethical implications of a technology. While Palm and Hansson note that feedback from a variety of “relevant stakeholders” outside of a company should be included, they do not provide a clear or practical means for acquiring this feedback. The lack of a broad source of feedback can result in narrow sources of information that can skew the results of an ethical analysis, burying concerns that are clear to an external relevant stakeholder group but are unclear to or *prima facie* undervalued by the product developer (Schaper-Rinkel 2013; Lucivero et al. 2011).

To address this issue, Lucivero (2016) provides a modified version of eTA called *Ethics of Emerging Technologies* (EET) that looks at the expectations of feasibility, usability, and desirability of an emerging technology amongst various relevant social groups, including developers, regulators, users, media, and other members of an industry (Lucivero et al. 2011). Each may have differing expectations of a product, and these expectations can be gleaned by analysing marketing materials, spokespersons comments, the views of project managers, and assessments by regulators of similar products in the field. However, EET may run into similar issues of time costs and ambiguity about which ethical framework to use in a particular setting.

3.4 Far Future Techniques (Techno-Ethical Scenarios Approach, TES)

Pioneered by researchers in biotechnology ethics like Marianne Boenink, Tsjalling Swierstra, and Dirk Stermerding, this approach uses “scenario-analysis” to evaluate the “mutual interaction between technology and morality” and to account for how one may affect the other (Boenink et al. 2010). Unlike eTA, TES does not require an ongoing procedure of dialogue, though regular check-ins are encouraged. Like eTA, TES seeks to develop the most likely scenario instead of multiple possible outcomes. However, TES is more long-term focused than most other ethical forecasting strategies and seeks to identify areas of micro change that may in turn create meso and macro changes over time. As an example, a

TES practitioner might note that the development of self-driving cars may create micro-changes in urban planning that may affect the moral understanding of driving a petrol-powered car, which may in turn inspire meso-level changes in insurance policies and laws governing accountability for traffic accidents, thereby inspiring a macro-level change in moral accountability for negligence and death.

TES embraces the dynamism of ethics and society by approaching a forecast of the ethical implications of a specific technology in three stages: first, the forecaster begins by “sketching the moral landscape” to “delineate the subject and give...some idea about past and current controversies and how they were solved” (Boenink et al. 2010). This process provides the scope of the analysis and identifies the relevant historical context for identifying common ethical issues that similar technologies have encountered in the past. Second, the forecaster generates “potential moral controversies” using the NEST-(New and Emerging Science and Technology) ethics model, an analytical framework that uses three sub-steps to identify common patterns in ethical debates about emerging technologies:

- (a) Identifying the promises and expectations that the technology or artefact offers—what does it enable or disable?
- (b) Imagining critical objections raised against the plausibility of its promises—does the technology or artefact actually accomplish what it proposes to? Is it feasible?
- (c) Constructing patterns or chains of arguments based off these critical objections—what are the positive and negative side effects of this technology or artefact? How will the moral debate around this technology or artefact develop? What kinds of resolutions exist to these controversies? (Swierstra and Rip 2007).

NEST-ethics helps the forecaster estimate how the moral debate over a new technology or artefact may affect its development. As Brey (2012) notes, literature reviews of ethical issues and workshops with policymakers are good sources of material for this stage, and Lucivero (2016) adds public company and industry statements, surveys of potential users, and statements by members of policy organizations to that list. The end result of this second step is that the forecaster has generated a list of potential moral controversies and resolutions that are likely to arise from an emerging technology.

After generating this list, the final step in the TES methodology is to construct “closure by judging the plausibility of resolutions” (Boenink et al. 2010). Similar to the role of the facilitator in Delphi, the forecaster must narrow down which moral controversies and resolutions are most plausible based on an analysis of historical moral trends of a technology and analogous examples (Swierstra et al. 2010; Boenink et al. 2010). The end result of these three steps is to provide the forecaster with the material to craft potential scenarios of the ethical issues that an emerging technology or artefact may stir up. For example, Boenink and colleagues use this model to create a scenario for evaluating the moral debate over medical experimentation on human beings (Boenink et al. 2010). By evaluating existing ethical frameworks for human experimentation, the historical basis for these frameworks, and current advances in this field relating to stem cells and

data mining, Boenink and his team create a scenario of the kinds of moral controversies that may arise in the medical sector.

3.4.1 Evaluation

TES provides numerous benefits over other ethical foresight models in that it is more focused on predictions for the distant future and it accommodates the ways in which technology and morality may influence each other. However, TES suffers from several weaknesses as a model to predict future ethical issues. First, it is useful for describing potential ethical issues and resolutions that *can* occur, but not at prescribing what ethical issues and resolutions *ought* to occur given the technology at hand. TES also struggles to identify which moral controversies are *most pressing* to the technology at hand (Brey 2012). Additionally, looking too closely at past historical and analogous ethical issues may result in new ethical concerns being missed entirely, particularly if they relate to relevant social groups that are not represented in previous literature. This approach, therefore, may pose weaknesses for cutting-edge technologies that have few analogous cases to draw upon for guidance.

3.5 Government and Policy Planning Techniques (ETICA)

ETICA (ETHical Issues of emerging iCt Applications) is a form of foresight analysis used by organizations like the European Commission (Floridi 2014) to guide policy-making decisions. At its core, ETICA identifies multiple possible futures that a particular technology may take and maps its social impacts (Stahl and Flick 2011). Unlike other foresight methods, ETICA explicitly distinguishes between the features that a technology has, the artefacts that encompass those features, and the application of those artefacts in its analysis (Stahl et al. 2013).

The first stage of ETICA involves the identification of known ethical issues with a particular emerging technology. ETICA relies on various methods to achieve this purpose, including bibliometric analysis of existing literature from governments, academics, and public reports on a particular technology to identify common ethical concerns (European Commission Research and Innovation Policy 2011). This source data is then used to create “a range of projected artefacts and applications for particular emerging technologies, along with capabilities, constraints and social impacts” (Brey 2012). For example, a bibliometric analysis of ICT literature, funded by the EU Commission in 2011, identified numerous ethical issues such as privacy, autonomy, data protection, and digital divides in these areas (Directorate General for Internal Policies 2011). Another method used in ETICA is written thought experiments to flesh out the ethical issues with an emerging technology. For example, Stahl (2013) uses a science-fiction like short story of the decision of a virtual intelligence to annihilate itself to identify ethical issues relating to identity and agency in the field of artificial intelligence.

After ethical issues are collected, they are then prioritized and ordered in an evaluation stage. For example, an analysis of robotics identified several particular issues with robotic applications such as human-like robotic systems and behavioural

autonomy (European Commission Research and Innovation Policy 2011). In the second evaluation stage of ICTs, the working group analysed these ethical issues according to four different perspectives: law, gender, institutional ethics, and technology assessment. They then compared and ordered common issues within ICTs, such as ambient intelligence, augmented and virtual reality, future internet, and robotics and artificial intelligence. This ranking is then used in the third stage of the ETICA process to formulate priorities for governance recommendations for policy-makers. These recommendations include setting up stakeholder forums to provide feedback on the process and incorporating ethics into ICT research and development (European Commission Research and Innovation Policy 2011; Stahl et al. 2013).

3.5.1 Evaluation

ETICA has numerous drawbacks that are similar to TES and eTA, most notably the fact that its sources for assessments of the major ethical concerns of a technology are not individually scrutinized and verified. A second major concern is the lack of specific actionable insights from such a foresight strategy. Recommendations tend to focus on broad principles, such as realizing “that ethical issues are context-dependent and need specific attention of individuals with local knowledge and understanding” (Directorate General for Internal Policies 2011). While this process is helpful for identifying a starting point for future discussions of ethical principles, it does not appear useful for providing clear actionable feedback on how to evaluate their technologies.

3.6 Combinatory Techniques (Anticipatory Technology Ethics, ATE)

Pioneered by Philip Brey and Deborah Johnson, *Anticipatory Technology Ethics* (ATE) evaluates technologies during the design phase as a means of reflecting how a ‘technology’s affordances will impact their use and potential consequence’ (Shilton 2015). ATE helps design teams consider the ethical values they bake into a technology by forecasting the kinds of issues that can arise at three different levels: an entire technology, artefacts of that technology, and applications of those artefacts by different relevant social groups (Brey 2012). ATE is founded on the principle that a robust form of EFA must analyse all three levels at once in order to provide a holistic understanding of a particular technology or artefact’s impact on society, thereby reducing ethical uncertainty. At the technology and artefact levels, ethical issues may arise due to inherent characteristics of the technology or artefact, unavoidable outcomes of its development, or certain applications that create such extreme moral controversy that the technology or artefact itself is morally suspect. As an example of this analysis at the technology level, the proliferation and development of nuclear technology enables the weaponization of that technology into artefacts that are so harmful they may justify ceasing development of any nuclear technology entirely. An example of this analysis at the artefact level is the ethical questions that arise from petrol-powered automobiles and their effects on the environment. In the latter analysis, the focus is not on the ethics of an entire technology (automotive) but

simply on the development of a specific artefact (petrol-powered vehicles as opposed to electric or solar-powered vehicles).

Like ETICA, ATE uses an identification stage to flag ethical issues and an evaluation stage to analyse which issues are likely to arise and how we should prioritize them. Brey recommends using an ethics checklist during the identification stage as a means to cross-reference the kinds of ethical issues that are likely to arise. In the application level, the ethicist's focus shifts to how different relevant social groups create a particular artefact's "context of use" (Brey 2012). Here Brey differentiates between *intended uses*, *unintended uses*, and *collateral effects* of an application of a technology as a basis for analysing the ethical issues at play. Intended uses represent what developers of an emerging technology and relevant social groups understand the purpose of that technology to be at the development and introduction stage. For example, powerful video editing software is intended for use by creative professionals working in the film industry. Unintended uses, however, might include the use of that same technology by terrorist organizations to develop high-quality propaganda videos. An example of possible collateral outcomes would be if this new video editing software, which requires high levels of expertise and resources to master, becomes the status quo for the film industry, thereby making it more difficult and expensive for aspiring low-income film producers to break into the industry.

The next step in ATE is the development of clear ethical issues at each level of analysis. At the technology level, Brey recommends using interviews and focus groups with experts in a particular subject matter to flesh out the relevant ethical issues a technology may raise. Mirroring the practice of eTA, practitioners should prioritize consulting experts in a particular technology field as they are best placed to understand the plethora of ethical issues at stake. Brey also argues that most ethical issues arise at the artefact and application levels in any case (Brey 2012). An example of ethical foresight at this level is Grace et al. (2017)'s paper examining the likelihood of an artificial intelligence-caused extinction level event by surveying experts in AI and machine learning.

At the artefact and application level, the use of existing foresight methodologies like ETICA and TES help provide a clear understanding of the kinds of artefacts that may develop. Surveys and interviews are common methods at these levels to identify future artefacts and applications. This analysis would include forecasts of how an artefact or application is likely to evolve over time and how that artefact or application will combine with others to create new kinds of artefacts. The goal of this analysis in ATE is to develop multiple possible futures that are plausible, as opposed to one single prediction (as in the case of Delphi or prediction markets).

3.6.1 Evaluation

ATE attempts to combine several of the strongest features of other EFA methodologies above. However, it suffers from similar weaknesses in terms of the time cost of an ongoing process and a lack of a clear conception of how to source relevant material for its analyses. In particular, checklists may be incomplete or may differ wildly depending on the kind of ethical framework one is starting with. Depending on the values of the society in which the technology is launched, an ethicist may wish to

use a different moral framework to develop and interpret the ethical issues that may arise (e.g., starting from a framework of conservative values, European values, or Christian values may result in different elements on a checklist or interpretations of those elements).

Shilton (2015) also notes several challenges with implementing ATE in certain situations. First, ATE's effectiveness is diminished if the technology in question is 'infrastructural' rather than user-facing as design teams may struggle to imagine an infrastructural tool in a social context with clear norms and rules. Infrastructural technologies also tend to have 'features of their design that frustrate a number of the techniques that might ordinarily encourage engineers to consider social contexts and social implications,' thereby creating more 'ethical outs' for engineers to argue that a particular issue should be resolved by other actors (e.g. government or the public). Lastly, Shilton notes that design teams usually need a prototype of a technology finished before they can begin an evaluation, which often means that certain design values are concretized into a technology.

4 Discussion: Known Limitations of EFA

Given the previous analysis of the relevant, current EFA methodologies, we can now evaluate some of their limitations and weaknesses as they are applied within technology companies. One core limitation relates to a lack of clarity over *who*, within a technology firm, should be conducting ethical foresight, and *where* the review process should sit within a firm. Most EFA methods provide a method for conducting an inquiry. However, with a few notable exceptions, they fail to provide more practical guidance about the ideal makeup of the foresight team or guidance on how to scale-up and implement this process in a large organisation.

A second limitation relates to a fairly basic point: what ethical framework should a company start with, in order to determine what ethical issues are of greatest concern? This is particularly true in cases where a company is developing transnational products that span across multiple cultures, legal systems, languages, and socio-economic demographics. Some communities, for example, may end up placing a greater value on privacy or personal autonomy than others, which can complicate the process of prioritizing or interpreting which ethical analysis may be preferable. Problem formulation is a key challenge with most EFA methods that can be made even more challenging in the context of products and services launched globally.

It is difficult to prescribe a specific ethical framework for evaluating ethical issues that a technology raises. Checklist impact assessments, like Privacy Impact Assessments and Algorithmic Impact Assessments, are informed by legal codes or industry-developed standards, but ethical codes are often more fluid, varied depending on one's moral framework, and difficult to interpret. Different moral frameworks may come to significantly different outcomes for the same issue, which is why EFA methodologies start from a "theory independent" stance that does not prescribe a particular framework over another (Grunwald 2000). Deciding which ethical framework to apply in a foresight analysis is a reasoned preference to be determined by the ethicist and not something that a foresight analysis can determine for an ethicist

a priori, irrespective of empirical evidence and historical circumstances. Recommendations for which ethical framework to use are thus out of scope in this article. Some ethical frameworks, such as eTA, encourage the ethicist to apply multiple different ethical frameworks in an analysis. This is a promising recommendation and one that we see as a viable starting point for a future EFA methodology. Another viable alternative is to adopt a regional ethical framework that may work as a benchmark, like the “Ethics Guidelines for Trustworthy AI” published in April 2019 by the European Commission’s High Level Group on Artificial Intelligence.

A third limitation of current EFA methods is their narrow focus on forecasting ethical issues of a particular technology that are agnostic to the organisation developing that technology. Major technology firms carry significant reputational and political power and often maintain a portfolio of technologies that intersect with each other in a variety of ways (Zingales 2017). For example, Facebook’s decision to launch a new dating app raised ethical issues beyond those of any other dating app company, namely relating to data usage and how this service would interact with other Facebook products (Statt 2018). Current EFA methods do not adequately account for this kind of bespoke organisational consideration.

A final limitation of current EFA methods is their reliance on qualitative methods of interviewing and focus groups to identify ethical issues. Private companies are often disincentivized to use these techniques in an assessment of their own technologies due to significant time and resource costs and potential risks to proprietary information being publicly released. Instead, firms are incentivized to develop and deploy technologies as quickly as possible to maximize their value, particularly in cases of technological ‘first movers.’ These incentives are at odds with current EFA approaches that require extensive amounts of time, openness, and self-reflection to succeed. It is our opinion that no effective EFA process can fully overcome this limitation, but we do offer some ideas below for ways to mitigate these concerns.

5 Recommendations for Potential Future Approaches to EFA

The limitations discussed above present several areas where future EFA methodologies can be improved. First, there is a clear need for the development of a new EFA methodology that is tailored to the specific needs of individual actors, no matter whether private or public, operating in the tech industry. Such a methodology should not only address common areas of concern between competing organisations but also how a particular organisation’s reputation, standing, or holistic product portfolio can affect the ethical issues that may arise in a single product. For example, if a major tech firm intends to launch a new social media app, how can that company forecast the ethical impacts of that app on different users? And how can that analysis consider the specific nature of that company, which may affect the kinds of ethical issues that arise? An ethical foresight analysis would ideally incorporate company-specific elements into its analysis, such as potential data sharing issues between an app and other products, or ethical issues that may arise from specific user demographics.

Second, future EFA methods should focus on ways to identify ethical risks at scale. There are promising developments in the fields of simulation and artificial intelligence that can complement or augment current qualitative methods of EFA, such as adversarial ‘red teaming’ of a technology to test for potential ethical weaknesses. For example, by using training data of known malicious actors, a firm could develop AI agents that simulate malicious users attacking that system (Floridi 2019). If the emerging technology in question were a dating app that contains many similar features to previous data applications, these agents could replicate various kinds of malicious behaviour and, in a staging environment, run simulated “attacks” to test the platform’s acceptable use policies. Currently, similar practices are in place in war games simulation exercises (Kania 2017; Knapp 2018) and in AI safety-robustness studies. In May of 2018, for example, a team at MIT trained a “psychopathic” image captioning AI using user posts from the social media website Reddit (Stephen 2018). By modelling a diverse array of users, a developer could test-run user interactions in a test environment version of their artefact, which may in turn identify ethical repercussions of their product’s affordances and features. There are clear limits to applying simulations to ethical foresight, the most significant being that a simulation requires a staged environment where one can determine the end criteria that a simulation is meant to meet. There may also be far too many variables and little enough data in most scenarios to create reliable simulations. However, given the pace of technological advancement in the field of Generative Adversarial Networks (GANs), this is one promising area where a new methodology for ethical foresight should be investigated for some situations.

Lastly, new EFA methods will only be effective if they are used as one part of a multi-pronged internal ethics process with strong executive-level support. EFA should be understood as one tool in the toolbox for internal ethics teams to use alongside other methods, including pre and post-deployment impact assessments, user focus groups, and other forms of user-centric design. As Latonero (2018) notes, human rights impact assessments are another valuable method approach that centers human rights at the heart of technology assessment. Other promising methods proposed by academic labs and technology firms include submitting anonymised case studies of challenging AI ethics issues to public consultation (Princeton Dialogue on AI Ethics 2018) or using public Requests for Information to source potential ethical concerns and mitigation options (Newton 2018). New EFA methods should ideally seek to complement and augment these other methods and processes.

6 Conclusion

This article has identified the current status of ethical foresight methodologies available today. It began by identifying the areas of study and related foresight methodologies in which ethical foresight is rooted, explaining what the purpose of ethical foresight is compared to other foresight methodologies. It then moved into a comparative analysis of several contemporary methodologies for ethical foresight to identify strengths and weaknesses. It included a description of potential future areas of research for ethical foresight methodologies and a discussion

that suggests there is significant value in developing a new type of Ethical Foresight Analysis that addresses some of the concerns of major corporate stakeholders that are developing emerging technologies and new artefacts. After reviewing existing ethical foresight methodologies, five common themes and criteria become apparent:

- (1) EFA methodologies cannot determine what ethical framework to use. The recommendation is that any EFA methodology should be “theory independent” and that practitioners should use multiple ethical theories in their analysis for a more robust final product. More research will be needed to identify minimal criteria of ethical clearance and ways in which such criteria may be made more stringent depending on needs, requirements, and the availability of adoptable ethical frameworks.
- (2) EFA methodologies seek to forecast a combination of the ethical impacts of a technology, an artefact or product, and the application of that artefact and product. More research is needed to identify how the most robust kinds of current ethical foresight models can evaluate on all three levels of analysis to provide a holistic picture of the ethical challenges that will arise.
- (3) All EFA methodologies reviewed use ongoing qualitative methods of data gathering to construct expectations of the ethical issues that an emerging technology may cause. This process can use either consensus-building strategies to gather a wide source of viewpoints via conferences, written opinions, and other open forums, or focus solely on experts who are well versed in the specifics of a given technology. The latter seems to work best when analysing an entire technological area, whereas the former appear to be more appropriate in instances of product/artefact/application evaluations. Research in quantitative methods, including AI-based simulations, is a promising and increasingly feasible approach, but one that still needs to be investigated.
- (4) EFA methodologies are iterative and build off one another in order to inform the design process of a technology on an ongoing basis. Ethical foresight tends to fail when used as a “one off” event, given the constantly shifting variables that make “stagnant” analysis of ethical issues quickly outdated. A future project will need to study how to overcome such methodological brittleness.
- (5) EFA methodologies tend to split between reducing information down into a single predictive probability or set of guidelines vs generating multiple potential futures that may occur. Depending on how a company intends to use EFA, certain strategies may prove more effective than others. In this case too, a full methodology will need to design strategies for the evaluation of the right approach.

There is a clear need for a new methodology of ethical foresight, one that addresses the requisites of an individual organisation that wishes to analyse the potential ethical outcomes of their products or services. It is clear that this gap in EFA methodologies is one that can be filled by supplementing existing EFA methodologies with new methodological and theoretical components. In the digital sector, where innovation can lead to dire ethical consequences, the ability

to forecast more accurately these risks can provide a unique opportunity for the advancement of corporate and social responsibility.

Acknowledgements Research for this article has been supported by Privacy and Trust Stream—Social lead of the PETRAS Internet of Things research hub. PETRAS is funded by the Engineering and Physical Sciences Research Council (EPSRC), Grant Agreement No. EP/N023013/1. This research has also been supported by a Facebook academic Grant in 2018.

References

- Armstrong, J. S. (2008). Methods to elicit forecasts from groups: Delphi and prediction markets compared. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1153124>.
- Ausloos J., Heyman R., Bertels N., Pierson J., & Valcke P. (2018) Designing-by-debate: A blueprint for responsible data-driven research & innovation. In: Responsible research and innovation actions in science education, gender and ethics. Springer Briefs in Research and Innovation Governance. https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1711945&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1. Accessed 9 Jan 2020.
- Battisti, D. (2004). The Italian way to e-democracy. A new agenda for e-democracy—position paper for an OII Symposium. Oxford Internet Institute. https://www.oii.ox.ac.uk/archive/downloads/publications/OIIPP_20040506-eDemocracy_200408.pdf.
- Bell, W. (2017). *Foundations of futures studies: Volume 1: History, purposes, and knowledge: Human science for a new era*. Abingdon: Routledge.
- Boenink, M., Swierstra, T., & Stemerding, D. (2010). Anticipating the interaction between technology and morality: A scenario study of experimenting with humans in Bionanotechnology. *Studies in Ethics, Law, and Technology*, 4(2), 1–38. <https://doi.org/10.2202/1941-6008.1098>.
- Braun, E. (1998). *Technology in context: Technology assessment for managers*. New York: Routledge.
- Brey, P. A. E. (2012). Anticipatory ethics for emerging technologies. *NanoEthics*, 6(1), 1–13. <https://doi.org/10.1007/s11569-012-0141-7>.
- Cantin, R., & Michel, P. (2003). Towards a new technology future approach. *Futures*, 35(3), 189–201. [https://doi.org/10.1016/S0016-3287\(02\)00063-0](https://doi.org/10.1016/S0016-3287(02)00063-0).
- Chansanchai, T. (2014). ‘A great laboratory for researchers’ launches with the Microsoft Prediction Lab. <https://blogs.microsoft.com/ai/great-laboratory-researchers-launches-microsoft-prediction-lab/>. Accessed 9 Jan 2020.
- Clarke, R. (2009). Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2), 123–135. <https://doi.org/10.1016/j.clsr.2009.02.002>.
- Collins, H. M. (1981). Stages in the empirical programme of relativism. *Social Studies of Science*, 11(1), 3–10. <https://doi.org/10.1177/030631278101100101>.
- Cowgill, B. (2005). Putting crowd wisdom to work. <https://googleblog.blogspot.com/2005/09/putting-crowd-wisdom-to-work.html>. Accessed 9 Jan 2020.
- Diffenbach, J. (1981). A compatibility approach to scenario evaluation. *Technological Forecasting and Social Change*, 19(2), 161–174. [https://doi.org/10.1016/0040-1625\(81\)90013-5](https://doi.org/10.1016/0040-1625(81)90013-5).
- Directorate General for Internal Policies. (2011). Pathways towards responsible ICT innovation—Policy brief of STOA on the ETICA project. European Commission.

-
- European Commission Research and Innovation Policy. (2011). Towards Responsible Research and innovation in the Information and Communication Technologies and Security Technologies Fields. A report from the European Commission Services.
- Fischer, C. S. (1992). *America calling: A social history of the telephone to 1940*. Berkeley: University of California Press.
- Floridi, L. (2014). Technoscience and ethics foresight. *Philosophy & Technology*, 27(4), 499–501. <https://doi.org/10.1007/s13347-014-0180-9>.
- Floridi, L. (2019). What the near future of artificial intelligence could be. *Philosophy & Technology*, 32(1), 1–15. <https://doi.org/10.1007/s13347-019-00345-y>.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When will AI exceed human performance? Evidence from AI experts. ArXiv:1705.08807 [Cs]. Retrieved from <http://arxiv.org/abs/1705.08807>.
- Grunwald, A. (2000). Against over-estimating the role of ethics in technology development. *Science and Engineering Ethics*, 6(2), 181–196.
- Guston, D., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society*, 24, 93–109.
- Jasanoff, S. (Ed.). (2010). *States of knowledge: the co-production of science and social order (transferred to digital print)*. London: Routledge.
- Johnson, D. G. (2007). Ethics and technology “in the making”: An essay on the challenge of nanoethics. *NanoEthics*, 1(1), 21–30.
- Kania, E. (2017). China’s quest for an AI revolution in warfare. The strategy bridge. <https://thestrategybridge.org/the-bridge/2017/6/8/-chinas-quest-for-an-ai-revolution-in-warfare>. Accessed 9 Jan 2020.
- Keller, P., & Ledergerber, U. (1998). Bimodal system dynamic a technology assessment and forecasting approach. *Technological Forecasting and Social Change*, 58(1–2), 47–52. [https://doi.org/10.1016/S0040-1625\(97\)00054-1](https://doi.org/10.1016/S0040-1625(97)00054-1).
- Knapp, B. (2018). Here’s where the Pentagon wants to invest in artificial intelligence in 2019. Defense News. <http://www.defensenews.com/intel-geoint/2018/02/16/heres-where-the-pentagon-wants-to-invest-in-artificial-intelligence-in-2019/>. Accessed 9 Jan 2020.
- Latonero, M. (2018). Governing artificial intelligence: Upholding human rights & dignity. *Data & Society*. <https://datasociety.net/output/governing-artificial-intelligence/>. Accessed 9 Jan 2020.
- Lucivero, F. (2016). *Ethical assessments of emerging technologies*. Cham: Springer. https://doi.org/10.1007/978-3-319-23282-9_8.
- Lucivero, F., Swierstra, T., & Boenink, M. (2011). Assessing expectations: Towards a toolbox for an ethics of emerging technologies. *NanoEthics*, 5(2), 129. <https://doi.org/10.1007/s11569-011-0119-x>.
- Manley, R. A. (2013). The Policy Delphi: A method for identifying intended and unintended consequences of educational policy. *Policy Futures in Education*, 11(6), 755–768. <https://doi.org/10.2304/pfie.2013.11.6.755>.
- Miles, I. (2010). The development of technology foresight: A review. *Technological Forecasting and Social Change*, 77(9), 1448–1456. <https://doi.org/10.1016/j.techfore.2010.07.016>.
- Millar, K., Thorstensen, E., Tomkins, S., et al. (2007). Developing the ethical Delphi. *Journal of Agricultural and Environmental Ethics*, 20, 53. <https://doi.org/10.1007/s10806-006-9022-9>.
- Mulkay, M. (1979). Knowledge and utility: Implications for the sociology of knowledge. *Social Studies of Science*, 9(1), 63–80. <https://doi.org/10.1177/030631277900900103>.
- Nazarko, L. (2017). Future-oriented technology assessment. *Procedia Engineering*, 182, 504–509. <https://doi.org/10.1016/j.proeng.2017.03.144>.
- Newton, C. (2018). Facebook’s Supreme Court for content moderation is coming into focus. *The Verge*. <https://www.theverge.com/interface/2019/6/28/18761357/facebook-independent-oversight-board-report-zuckerberg>. Accessed 9 Jan 2020.
- Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change*, 73(5), 543–558. <https://doi.org/10.1016/j.techfore.2005.06.002>.
- Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, 14(3), 399–441. <https://doi.org/10.1177/030631284014003004>.
- Polgreen, P. M., Nelson, F. D., Neumann, G. R., & Weinstein, R. A. (2007). Use of prediction markets to forecast infectious disease activity. *Clinical Infectious Diseases*, 44(2), 272–279. <https://doi.org/10.1086/510427>.

-
- Princeton dialogue on AI ethics. (2018). <https://aiethics.princeton.edu/>.
- Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic Impact Assessments, 22. <https://ainowinstitute.org/aiareport2018.pdf>.
- Rip, A., Schot, J., & Misa, T. J. (1995). *Managing technology in society: the approach of constructive technology assessment*. London: Pinter Publishers.
- Schaper-Rinkel, P. (2013). The role of future-oriented technology analysis in the governance of emerging technologies: The example of nanotechnology. *Technological Forecasting and Social Change*, 80(3), 444–452. <https://doi.org/10.1016/j.techfore.2012.10.007>.
- Shilton, K. (2015). “That’s not an architecture problem!”: Techniques and challenges for practicing anticipatory technology ethics. iConference 2015 Proceedings, <http://hdl.handle.net/2142/73672>.
- Stahl, B. C. (2013). Virtual suicide and other ethical issues of emerging information technologies. *Futures*, 50, 35–43. <https://doi.org/10.1016/j.futures.2013.03.004>.
- Stahl, B. C., & Flick, C. (2011). ETICA workshop on computer ethics: Exploring normative issues. In S. Fischer-Hübner, P. Duquenoy, M. Hansen, R. Leenes, & G. Zhang (Eds.), *Privacy and identity management for life* (Vol. 352, pp. 64–77). Berlin: Springer. https://doi.org/10.1007/978-3-642-20769-3_6.
- Stahl, B., Jirotko, M., Eden, G., Computing, C. F., & Responsibility, S. (2013). Responsible research and innovation in information and communication technology: Identifying and engaging with the ethical implications of ICTs.
- Statt, N. (2018). Facebook is taking on Tinder with new dating features—The Verge. *The Verge*. <https://www.theverge.com/2018/5/1/17307782/facebook-tinder-dating-app-f8-match-ocupid>. Accessed 9 Jan 2020.
- Stephen, B. (2018). MIT fed an AI data from Reddit, and now it thinks of nothing but murder. Retrieved June 8, 2018, from <https://www.theverge.com/2018/6/7/17437454/mit-ai-psychopathic-reddit-data-algorithmic-bias>. Accessed 9 Jan 2020.
- Swierstra, T., Bovenkamp, H., & Trappenburg, M. (2010). Forging a fit between technology and morality: The Dutch debate on organ transplants. *Technology in Society*, 32, 55–64. <https://doi.org/10.1016/j.techsoc.2010.01.001>.
- Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of Moral argumentation about new and emerging science and technology. *NanoEthics*, 1(1), 3–20. <https://doi.org/10.1007/s11569-007-0005-8>.
- Tapio, P. (2003). Disaggregative policy Delphi: Using cluster analysis as a tool for systematic scenario formation. *Technological Forecasting and Social Change*, 70(1), 83–101. [https://doi.org/10.1016/S0040-1625\(01\)00177-9](https://doi.org/10.1016/S0040-1625(01)00177-9).
- Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. *Technological Forecasting and Social Change*, 75(9), 1396–1405.
- Turoff, M., & Linstone, H. A. (1975). *The Delphi method: Techniques and applications*. Boston: Addison-Wesley Pub. Co., Advanced Book Program.
- Van Eijndhoven, J. C. M. (1997). Technology assessment: Product or process? *Technological Forecasting and Social Change*, 54(2–3), 269–286. [https://doi.org/10.1016/S0040-1625\(96\)00210-7](https://doi.org/10.1016/S0040-1625(96)00210-7).
- van Est, R., & Brom, F. W. A. (2012). Technology assessment: Analytic and democratic practice. In R. Chadwick (Ed.), *Encyclopedia of Applied Ethics* (2nd ed., pp. 306–320). San Diego: Academic Press. <https://doi.org/10.1016/b978-0-12-373932-2.00010-7>.
- Zingales, L. (2017). Towards a political theory of the firm. *Journal of Economic Perspectives*, 31(3), 113–130.