



Anything new under the sun? Insights from a history of institutionalized AI ethics

Simone Casiraghi¹

Published online: 24 April 2023
© The Author(s) 2023

Abstract

Scholars, policymakers and organizations in the EU, especially at the level of the European Commission, have turned their attention to the ethics of (trustworthy and human-centric) Artificial Intelligence (AI). However, there has been little reflexivity on (1) the history of the ethics of AI as an institutionalized phenomenon and (2) the comparison to similar episodes of “ethification” in other fields, to highlight common (unresolved) challenges.

Contrary to some mainstream narratives, which stress how the increasing attention to ethical aspects of AI is due to the fast pace and increasing risks of technological developments, Science and Technology Studies(STS)-informed perspectives highlight that the rise of institutionalized assessment methods indicates a need for governments to gain more control of scientific research and to bring EU institutions closer to the public on controversies related to emerging technologies.

This article analyzes how different approaches of the recent past (i.e. bioethics, technology assessment (TA) and ethical, legal and social (ELS) research, Responsible Research and Innovation (RRI)) followed one another, often “in the name of ethics”, to address previous criticisms and/or to legitimate certain scientific and technological research programs. The focus is on how a brief history of the institutionalization of these approaches can provide insights into present challenges to the ethics of AI related to methodological issues, mobilization of expertise and public participation.

Keywords Institutionalization · Bioethics · Technology Assessment · Artificial Intelligence · European Commission · Responsible Research and Innovation · Public participation · Science and Technology Studies

Introduction

Especially in the last decade, several scholars, policymakers and organizations in the European Union (EU) have turned their attention to the ethics of (trustworthy and human-centric) Artificial Intelligence (AI). So far, scholarly works have gone in two main directions. On the one hand, some scholars have focused on which ethical principles are the most suited for this field and how they should be operationalized in practice (Jobin et al., 2019; Sartor, 2020), often looking at the precedent of bioethics (Floridi et al., 2018; Morley et al., 2021). On the other hand, other scholars have been skeptical of these invocations of AI ethics for several reasons, from

ethics washing (Wagner, 2018) to problems of representation or checks and balances (Delacroix & Wagner, 2021; Nemitz, 2018; van Dijk et al., 2021; Yeung et al., 2020). By contrast, there has been little reflexivity on (1) the history of the ethics of AI as an institutionalized phenomenon¹ and (2) the comparison to similar episodes of “ethification”² in other fields, to highlight common (unresolved) challenges.³

For the scope of this article, institutionalization refers to the creation of bodies, expert groups or committees with administrative and consultive functions, as a method for

✉ Simone Casiraghi
simone.casiraghi@vub.be

¹ Faculty of Law & Criminology, Law, Science, Technology and Society (LSTS) research group, Vrije Universiteit Brussel (VUB), Pleinlaan 2, Brussels 1050, Belgium

¹ This is not the case, for instance, in the field of Responsible Research and Innovation (RRI), which will be discussed below. See (Owen, Schomberg, et al., 2021; Shanley, 2021).

² Ethification is “the proliferation of invocations of ‘ethics’ and ‘ethical principles and values’ in legal, policy, academic and corporate discourses about the governance of technology and the growing institutional importance of ethical expertise through bodies such as ethical committees, advisory groups and boards, as well as through procedures enshrined in methodological documents such as ethical guidelines and checklists”. See (van Dijk et al., 2021, p. 2).

³ With some notable exceptions, e.g. (Mittelstadt, 2019)

decision-making in controversial issues involving science and technology. The focus is on institutionalization in more “governmental” terms, looking at initiatives of some prominent EU institutions and expert groups.⁴ In these contexts, ethics has had an increasing influence on the governance and regulation of emerging technologies at the EU level. The decisions, opinions and guidelines of ethics expert bodies have had consequences within the landscape of innovation governance. This warrants a critical check, both regarding the institutional aspects and the narratives about the history of the institutionalization process of ethics.

The dominant narrative on AI ethics, backed up not only by some scholars, but also by political institutions such as the European Commission (EC), is that quick digital technological developments are to some extent inevitable and/or necessary to address societal challenges, since they bring about terrific opportunities, but they also pose new unpredictable risks. Therefore, a mix of *ex-ante* and *ex-post* approaches, often, but not always, grouped under the label of “ethics”,⁵ that consider the most desirable paths for technological developments and assess their social impacts, are deemed necessary to govern emerging technologies and mitigate risks and harms to individuals. Many of these approaches have become, throughout the years and to different degrees and in different ways, institutionalized.

In the field of ethics of AI, this narrative can be observed in the European landscape in the documents and initiatives of the European Commission’s Robotics and AI Unit at the Directorate-General for Communications Networks, Content and Technology (DG Connect) (European Commission, 2018b, 2018a) and European Parliament’s Committee on Legal Affairs (JURI) on AI (García del Blanco, 2020), but also in the work of ethical advisory bodies like the European Group on Ethics in Science and new technologies (EGE) (European Group on Ethics in Science and New Technologies, 2018) or of expert groups created ad hoc to address ethical issues around the digital domain or AI, such as the EDPS EAG (European Data Protection Supervisor Ethics Advisory Group, 2018) or the AI HLEG (High-Level Expert Group on Artificial Intelligence, 2019a).

⁴ There are different ways to conceive institutionalization (Tallacchini, 2009, p. 287). While the focus of this article is on policy instruments for science and technology governance and development, institutionalization can also be conceived more broadly as the creation of university curricula, conferences and events, or of codes of conduct, in areas other than science and technology. Focusing on these alternative modes of institutionalization requires, however, separate work. See e.g. (Owen, Pansera, et al., 2021)

⁵ These methods have been labeled “ethical” even if they belong to different methodological and disciplinary traditions, including e.g. philosophy, sociology, management, media studies or computer science. See (Reijers et al., 2018).

However, this narrative seems to overlook certain aspects of the history of the institutionalization of ethics both in the US and the EU. Science and Technology Studies (STS)-informed perspectives have highlighted additional explanations for the rise of institutionalized ethics. Although there is no agreement on historical and sociological accounts of institutionalized ethics, since these phenomena are all still too recent (Eckenwiler & Cohn, 2007), STS scholars have described the ways institutions and organizations have long resorted to ethics in other fields, primarily in the life sciences (Jasanoff, 2011; Tallacchini, 2009), by looking at what the actors discursively perform and construct as ethics and by analysing the institutional settings in which different interests and strategies are pursued (van Dijk et al., 2021). Examples are the need for governments to gain more control of scientific research (e.g. bioethics or Technology Assessment (TA) in the US) or to bring institutions closer to the public on controversies related to emerging technologies (e.g. in the EU, Jecker et al., 1997).

Building on these insights, this article analyzes how different approaches, i.e., bioethics, TA, Ethical Legal and Social (ELS) research and Responsible Research and Innovation (RRI), followed one another, often “in the name of ethics”, to address previous criticisms and/or to legitimate certain scientific and technological research programs. It focuses on how some historical and sociological accounts (Evans, 2012; Jecker et al., 1997; Jonsen, 1998; Rothman, 1991) of these approaches can provide insights into present challenges to the ethics of AI related to methodological issues, mobilization of expertise and public participation. This brief history does not only highlight disruptive changes but also continuity, i.e. how, in practice and partially in contrast with the proclaims and visions of some programmatic documents, different ethical approaches and the ethics of AI⁶ still co-exist and draw on each other.

The structure⁷ of the article is the following. In Section 2, an overview of traditional bioethics in the US will be presented,⁸ starting from its origin in the 1960s. The political and methodological context around its birth will be sketched and compared to that of Europe in the early 1990s. Section 3 will summarize the history of TA, looking into the rise and fall of the Office of Technology Assessment (OTA) in the US and into the several initiatives that

⁶ Since several articles have been already written mapping the current institutionalization of AI ethics, this part will be omitted here due to space constraints. See e.g. (Hagendorff, 2020; Jobin et al., 2019; van Dijk et al., 2021)

⁷ I thank the anonymous reviewer for the useful suggestions to restructure part of the paper.

⁸ The scope of the article is the EU, but, since the US influenced the EU in the development of bioethics and other forms of ethics, it is important to look at the American context too before moving to the EU.

took inspiration from it in Europe since the 1980s. Section 4 will sketch the spread of ELS- programs in the US and EU. Each Section will end with a subsection (2.3, 3.3 and 4.3) in which “older” forms of institutionalized ethics are compared to the current AI ethics. Section 5 will conclude and provide some recommendations.

Common morality and “principlism” bioethics

The origins in the US: public scandals and the birth of bioethics

The term “bioethics” was revived⁹ in the US by biochemist Van Rensselaer Potter (1911–2001) as an ethical reflection on human, animal and natural life triggered by scientific and technological developments (Potter, 1970). In the same years, it is André Hellegers, physician and theology professor (1926–1979), and later the Georgetown model of bioethics,¹⁰ however, that conceived bioethics in a narrower sense, more related to “medical ethics”.¹¹

Bioethics in this narrow sense provided a new language that could address the problems resulting from a series of medical research scandals in the 1960s, when it was more and more questioned whether scientists should have the authority or decisional power over the ethics of their experiments. After public scandals appeared in the press¹² and discussions took place in the Senate, the US government adopted the National Research Act, which established in 1974 “The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research” (which lasted until 1978; hereinafter “the National Commission”).¹³

The National Commission published the Belmont Report and the Institutional Review Board (IRB) Report. The IRB report endorsed the creation of IRBs, independent ethical

committees to peer review publicly funded research,¹⁴ later legally established in 1991, the Federal Policy for the Protection of Human Subjects, also known as the “Common Rule”.¹⁵ The Belmont report, published in 1979, identified three basic principles for human research that were later given a scholarly/philosophical foundation, i.e. respect for persons (linked to informed consent), beneficence (linked to risk-benefit analysis) and justice.¹⁶ Such principles were not groundbreaking or new for scientists, but rather “a post hoc philosophical backfilling of justifications for practices scientists had supposedly endorsed for many years” (Faden & Beauchamp, 1986, p. 216).

Philosophical thought, as a form of secularized thinking (as opposed to theology), but still “external” and critical to science, eventually influenced the National Commission and a particular philosophy provided a background for the new profession of bioethics. Philosophers Tom Beauchamp and James Childress co-authored a textbook (Beauchamp & Childress, 2019) that formalized the principles of the Belmont report in a coherent philosophical theory, later called also “principlism”.

Also referred to as the common morality principle, this method has become the reference point for bioethics professionals, and not primarily for its academic excellence or internal coherence. The method, for instance, was endorsed by the US government, being attractive to government’s officials for the claim to represent common morality (Evans, 2012, p. 57) and speaking in the name of the values of the people. It was a good fit with American law (Evans, 2012, p. 48), having an appealing basic structure ensuring simplification of decision making as opposed to disorderly or somehow arbitrary systems used before, and with bureaucratic authorities in health care institutions. It provided an opportunity to increase trust, to counter the decline of trust in physicians, also due to the debates and scandals about human experiments and their coverage by the press (Rothman, 1991, p. 61). Finally, academically, it was a middle ground between the two competing moral theories of deontology and consequentialism and, in line with utilitarian

⁹ The term goes back to at least the German theologian Fritz Jahr (1895–1953). See (Jahr, 1927; Sass, 2007).

¹⁰ The first US-based institute for bioethics, the Kennedy Institute, was established at Georgetown University in Washington DC. See (Reich, 1995).

¹¹ This tradition goes back to the Greek physician Hippocrates (460–370 BC) but that was first coined by English physician Thomas Percival (1740–1804) in his book “Medical Ethics; or, a Code of Institutes and Precepts. Adapted to the Professional Conduct of Physicians and Surgeons” (1803).

¹² Especially the Tuskegee syphilis case, on which the New York Times published an article in 1972, available at <https://www.nytimes.com/1972/07/26/archives/syphilis-victims-in-us-study-went-untreated-for-40-years-syphilis.html>. When the story came out, the medical community ignored it or criticized the press for being irresponsible and interfering with research. See (McNeill, 1993, p. 61).

¹³ It was an independent body made of eleven members, five scientists and six nonscientists, including social scientists and theologians. See (Evans, 2012, p. 40; Yesley 2008, p. 3).

¹⁴ The concept of oversight by an independent committee was already introduced by the Declaration of Helsinki at Art. I.2: “The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol which should be transmitted to a specially appointed independent committee for consideration, comment and guidance.” (Text from the revision of 1975).

¹⁵ Department of Health and Human Services Rules and Regulations; Title 45; Code of Federal Regulations (CFR); Part 46.

¹⁶ The limited number of principles was chosen to simplify decision-making in bioethics and enhance calculability. A member of the National Commission, for instance, complained that in earlier drafts of the Belmont Report there were “too many principles” (originally 7) and that the list was not “crisp enough”. See (Evans, 2012, p. 52; Jonsen 1998, p. 103).

considerations (i.e. balancing risks and benefits), it fits with a pluralist society and market-oriented framing of regulatory issues on biotechnology (Jasanoff, 2005, p. 177).

The origins of bioethics in the EU (1991–1994)

Bioethics in the US was more a response to controversies questioning the authority of scientists, focusing on their individual responsibility because of big research scandals that attracted public attention. In the EU, by contrast, its history is related to solving the problem of democratic deficit and building a European identity (Jasanoff, 2005), in a moment of transition from the economic to the political unification of the European Community with the Maastricht Treaty of 1993 (Tallacchini, 2015, p. 164). In particular, the 1990s turn to ethics by the EC took place in a time of political clash between the vision of economic growth and competitive development held by the EC and the resistance or opposition of EU citizens and Member States (Busby et al., 2008, p. 804).¹⁷ In this historical context, ethics helped to mitigate the friction between the EC, EU citizens and EU Member States in two ways.

First, to bring EU institutions closer to the public,¹⁸ ethics offered a way out of the impasse, for example, between a clear Research & Development (R&D) agenda for further industrializing agriculture, and the risks to the environment perceived by citizens, while the official policy language downplayed value judgments on such risks, presenting them as “objective” science (Levidow & Carr, 1997). These problems were recognized in an EC communication in 1991 (European Commission, 1991), where ethics is associated with the need to avoid uncertainty and confusion in the public debate. In this communication, the EC sketched the role of what ethics should have been (European Commission, 1991, p. 16), i.e. (1) opening up the discussion to “interested parties” and (2) enabling experts to participate in the legislative process. Regarding the latter point, a few months later, in 1991, the Group of Advisers on the Ethical Implications of Biotechnology (GAEIB) was established, made of six members including scientists, lawyers, philosophers and theologians, to advise the EC and identify ethical issues raised by biotechnologies, as well as to inform the public.¹⁹

¹⁷ For instance, this was exemplified by debates concerning agricultural biotechnology, e.g. the “mad cow” crisis, Genetically Modified Organisms and hormone treated beef (Jasanoff, 2005, p. 89; Levidow & Carr 1997).

¹⁸ Ethics was used “as a social lubricant” to help increase the acceptability of technological changes that were supposed to serve the common interest (Levidow & Carr, 1997, p. 30).

¹⁹ The mandate of the GAEIB 1991–1997 is available at https://ec.europa.eu/archives/bepa/european-group-ethics/archive-mandates/mandate-1991-1997/index_en.htm.

Second, ethics allowed the EU legislator to address value-related issues in life sciences despite these being usually regulated at a national level, avoiding conflicts and negotiating between diverging positions of EU Member States.²⁰ The role of the GAEIB was thus also of political integration, making the discrepancies regarding moral visions at the EU and Member States’ level, as well as disharmony in the respective legal frameworks, coexist (Tallacchini, 2009, p. 293). One of the main contrasts in life sciences in the 1990s, for example, regarded the opposition to some forms of genetic engineering research, for instance in Germany, in contrast to more accepting and experimental attitudes, such as in the UK (Jasanoff, 2005).

The criticisms by the European Parliament²¹ contributed to the replacement of the GAEIB by the EGE in 1997, which, until today,²² has been tasked to cover all areas of application of science and technology, not just biotechnology. The list of opinions of GAEIB and EGE is a good example of how to “neutralize”²³ socially and politically divisive conflicts resulting from the EC policies in life sciences. The EGE became an “obligatory passage point” (Callon, 1986) for the EU legislator, since whenever directives had anything to do with values, their opinions had to be explicitly considered (Tallacchini, 2015, p. 165).

Bioethics and the limits of principle-based ethics

The system of traditional bioethics was successful throughout the 1970 and 1980 s in the US, but it started showing its major political weaknesses in the late 1980s, which were reflected in the EU in the 1990s. Despite its steady growing influence since the 1950s until today, in academia, hospitals and biomedical research centers, bioethics has been

²⁰ In this sense, ethics was used, paraphrasing (Levidow & Carr, 1997), also as a “political lubricant”. I thank the anonymous reviewer for this insightful remark.

²¹ The Parliament criticized the GAEIB for focusing too much on the interests of research and not enough on the possible effects on society; their sometimes-unclear statements; the fact that the EP was not consulted on the appointment of the members of the GAEIB. See (European Parliament, 1997). The GAEIB was also criticized earlier by Green MEPs: “Many public-interest groups see the Bioethics Committee as a tranquilizer pill for public opinion. [...] To instrumentalize the ethical debate in this way, by dictating *a priori* where its findings should lead, is an utterly unethical and unacceptable approach to ethics” in H. Breyer, “Committee on Energy, Research and Technology: Draft Response to Bangemann Report [CEC 1991],” 1992, 15.

²² Commission Decision (EU) 2021/156 of 9 February 2021 renewing the mandate of the European Group on Ethics in Science and New Technologies, C/2021/715, OJL 46, 10.2.2021, p. 34–39.

²³ The term, used by Jasanoff, refers to the instrumental use of ethical opinions to de-politicize, or technocratically defuse, value-laden conflicts (Jasanoff, 2005, p. 90).

said to be “in crisis” (Evans, 2012, p. 75; Solomon 2014),²⁴ due to the increasing (political) controversies about the legitimacy and authority of the field and its methodological orientations,²⁵ to the point of questioning its very existence (Turner, 2009).

Among these criticisms, scholars have noted a problem of “quality control” (Benatar, 2006). In a narrower sense, bioethics could be considered a branch of philosophical inquiry, having to meet the quality standards of academic philosophy. However, bioethics seems to be often interpreted in a much broader sense, with many practices involved under its umbrella, each one with its own terminology and methods. As a result, many people can speak in the name of bioethics but without the need to adhere to any constraint or peer-review process, thus leaving little room to check the soundness of an ethical argument, principle or reasoning. A parallel in the ethics of AI could be found in the confusion between fundamental rights, from a legal perspective, and ethical principles. In the AI HLEG’s guidelines, while the principles are said to be anchored in, and operationalization of, fundamental rights (High-Level Expert Group on Artificial Intelligence, 2019a, p. 10), the reduction of the latter to basic principles such as “prevention of harm” empty them out of their legal and normative meaning, thus leading to a “watering down” (Delacroix & Wagner, 2021, p. 4).

More specifically, due to its short life, in the field of AI ethics there is a lack of proven methods to translate principles into practice (Mittelstadt, 2019, p. 503). While bioethics, throughout the decades, has relied on more robust ways that medicine has developed to translate high-level principles into practical requirements (e.g., professional codes of conduct, ethics review committees or licensing schemes), the same cannot be said for AI yet. There is general agreement on high-level principles in the field (Jobin et al., 2019), but there is no accepted hierarchy of principles in case of conflicting norms as well as little enforcement and accountability mechanisms in place (Delacroix & Wagner, 2021, p. 7). However, several attempts are being made to translate principles into practice (European Commission, 2020; Morley et al., 2021), so one could argue that it is just a matter of time before principlism acquires the same robustness in AI as in the biomedical field.

²⁴ For Evans, this crisis regards mostly bioethics as a profession or “public policy bioethics” and not, for instance, research bioethics or healthcare ethics consultations (Evans, 2012, p. xxxiii).

²⁵ In a debate over healthcare reform in the US in 2009, for instance, conservatives used selected passages of some bioethicists on topics like euthanasia to erode support for the Democrats’ plan, to which famous bioethicists responded by defending their colleagues and more generally the role of bioethicists in society (Evans, 2012, p. xvi). Another controversy in the US was about the establishment of the so-called “Kass Council” (Briggle, 2009; Plomer, 2008).

Nonetheless, there are different problems with principlism as a method in the first place. Principlism, as an expression of analytic philosophy (Callahan, 1982, p. 4), has long been considered too narrow and simplifying compared to other richer approaches (Callahan, 1982). Since principlism became so dominant, groups of scholars who were more interested in *debating ends* rather than an abstract system of knowledge or “engineering philosophy” where ends were pre-defined, were cut off from the discussion (Evans, 2012, p. 65). All items about metaphysics, world view or the impacts of technologies on society at large became seen as beyond the scope of “mainstream” bioethics (Zwart et al., 2014, p. 6). Despite these criticisms, it is surprising to observe that AI ethics still heavily relies on the principle-based approach of bioethics. In analogy to bioethics, the ethical debate on AI seems largely limited to design and implementation, not on whether AI systems should be built in the first place (Greene et al., 2019, p. 2127). Ethical charters are framed as imperatives and not as discussions about the possibility of *not* doing anything. This was evident in the work of the AI HLEG, where the section discussed in the public draft of the ethical guidelines on “red lines” that should never be created by AI developers, was removed from the final version of the document (Metzinger, 2019). Many people in the group, it was reported, were not comfortable with the idea that more guidance was needed on distinguishing between what people *can* do about AI and what people *should* (not) do with AI systems (High-Level Expert Group on Artificial Intelligence, 2018, p. 11).

Technology assessment and anticipatory governance (1972–1995)

Origins: OTA in the US

Another major approach to understanding and assessing the impact of emerging technologies, Technology Assessment (Tallacchini, 2009, pp. 284–287), developed in the US in parallel with institutionalized bioethics. At least initially, narratives on ethics did not play an explicit role in the establishment of TA as a dominant science governance approach, but the development of institutionalized TA and (bio)ethics are interconnected. In common with bioethics, TA was initially expert-based, providing an alleged neutral and factual input, but afterward, because of criticisms, it broadened up to other stakeholders.²⁶ Additionally, it supported traditional approaches to legislation in governing science and technology (Paula, 2008, p. 1) and a form of cost-benefit analysis, to mitigate the costs of technological development and

²⁶ See Sect. “From ELSI and ELSA to RRI”.

maximize their benefits, in a kind of formal and transparent manner.

The Office of Technology Assessment (OTA) in the US was established by the US Congress in 1972, but its history has deeper origins in the 1960 and 1970s regarding (a) the concerns about the relations between technology and the environment and (b) the tensions between the executive and legislative branches of the US federal government (Bimber & Guston, 1997; Grunwald, 2009, p. 1104; Kunkle 1995, p. 176). As for b), in the light of technological developments, e.g. in aeronautics, Congress members felt they needed more and better technical advice to legislate and choose research and development policies on matters involving scientific and technical complexity, in contrast with the more thorough and wider scientific and technical advice the US president had (Kunkle, 1995, p. 177). As a result, the idea of creating a separate advisory mechanism for the Congress, independent from the executive branch, emerged.

Given the early suspicions and hostility on this vision of TA, e.g. that what it really meant was regulating technology (left-oriented “regulation in disguise” (Kunkle, 1995, p. 180)) or that it would infringe the authority of the Congress regarding policy-making powers, Emilio Daddario (1918–2010), and his successor as chairman John Davis (1916–1992), started emphasizing more the informative and supplementary role of a hypothetical Office of TA for the Congress, which would restore the imbalances of federal powers in the light of the advisory deficiencies of the Congress vs. executive agencies (Kunkle, 1995, p. 182).

In its early years, instead of a proactive channel for public input for a coherent technology policy, with a broad monitoring function, OTA was more of a means to provide technical advice, requested *ad hoc*, to members of the Congress. Its goal was more related to parity of information among executive and legislative branches rather than to assessing the consequences of technological developments. Outside experts, for instance, were consulted only *post hoc*, on topics that had already been decided beforehand by the Technology Assessment Board (TAB), i.e. the twelve members body governing the OTA (Kunkle, 1995, p. 188). In this first phase, TA mostly fulfilled an “early warning function”, by providing information on possible future “secondary effects” (i.e. social, cultural, political) of technologies to policymakers. However, it became soon clear that, in practice, impacts of technologies could only be partially foreseen and providing neutral information to policymakers was not possible. Therefore, OTA moved from an early warning function to the development of policy alternatives. OTA earned, until 1995, with this new approach, the reputation of a reliable, informative and “unbiased” source of information

for the Congress, producing hundreds of reports on a wide range of topics.²⁷

Just like for the case of bioethics, looking at the dynamics originating TA in the US is useful to better understand how it developed separately in Europe and in the EU’s and its Member States’ institutions.

TA in the EU

The OTA was an influential example for many European States that established their own agencies in the mid-1980s (Grunwald, 2019, p. 704; Van Eijndhoven 1997, p. 269).²⁸ Among the first ones there were the German Bundestag Technology Assessment Bureau (established in 1989), the Netherlands Office for Technology Assessment (established in 1986; later renamed Rathenau Institute), the French *Office Parlementaire d’Evaluation des Choix Scientifiques* (established in 1983), the UK Parliamentary Office of Science and Technology (established in 1989) and the Danish Board of Technology (established in 1986).²⁹ Together, the European TA institutions founded the European Parliamentary Technology Assessment (EPTA) network in 1990, later joined by other international countries.

The OTA also had an influence at a supra-national level, on the European Community. In 1978 the European Commission approved the Forecasting and Assessing for Science and Technology (FAST, 1978–1983)³⁰ to develop a coherent long-term policy on science and technology. The EP later established the Scientific Technology Options Assessment (STOA), originally launched as a pilot project in 1987, to provide objective, comprehensive and independent assessment of science and technology issues to the Parliament.³¹ One of the reasons behind the creation of STOA, just like OTA, was the idea to balance the power between the Parliament and EC on matters related to impacts of technologies (Van Eijndhoven, 1997, p. 273).

In its specific instances of Member States and EU institutions, TA evolved differently from the initial conception of

²⁷ The OTA online archive includes all the publications. Available at <https://ota.fas.org/otareports/>.

²⁸ Attempts to create a variant of the OTA in some European states, such as Germany and France, were already made in the 1970s (Smits & Leyden 1995 25). However, the proposals were turned down by the governments that feared that the parliaments would get too strong.

²⁹ Each of these organizations has its history and structure, reflecting cultural and political differences in each country (Van Eijndhoven, 1997, p. 271).

³⁰ Established by decision of the Council of the Ministers in July 1978. Its goal was to help Member States to choose objectives of future research and development options.

³¹ The STOA wrote some reports in the 1990s to advise the EC on the ethical aspects of EU-funded research. See e.g. (Chadwick et al., 1998, 1999)

OTA, both for political (e.g. differences in political systems)³² and contingent reasons (e.g. budget and capacity).³³ Two new features of (some) European TA stemmed from the idea that the decision-making process should be broadened. First, a strand of TA called “public” or “participatory” TA (Van Eijndhoven, 1997, p. 278), focused on empowering democracy and promoting a more pluralistic and inclusive approach (Van Eijndhoven, 1997, p. 278; Vig 1992, p. 5). Second, a strand, developed in the Netherlands, called “constructive” TA or CTA (Schot & Rip, 1997), had a more economically driven focus on new technological opportunities (Vig, 1992, p. 5), influencing technology development at early stages, where technologies are considered as evolving in close interaction with social systems, to select those technologies that can maximize social, economic and environmental benefits.

TA and the politicization of (ethical) expertise

Despite some similarities, the challenges faced by institutionalized TA in the US were different from those for bioethics. By the early 1990s, OTA had come to a dilemma: on the one hand, if it claimed more autonomy concerning technology policy, it would have been put down by the Republicans who, in the mid-1990s, controlled both houses of Congress (Kunkle, 1995, p. 193); on the other hand, if its role was only to provide ad hoc advice, it would have been seen as superfluous. OTA eventually closed in 1995, to reduce expenditures and avoid duplicate functions, and the OTA’s functions were relocated to the Library of Congress.³⁴

In parallel to the decline of OTA, however, but for different reasons and dynamics, TA in Europe progressively became a minor, marginalized tool in comparison to the “use” of ethics as an instrument to build a European identity or to promote market integration.³⁵ TA in the EU became a “residual instrument” for parliaments with a marginal position in the exploration of policy options, with little impact and too little use of results (Smits et al., 1995). Two major problems contributed to this outcome. A political problem was that the EU initiatives on TA had no respectable position due to a lack of political legitimation, little insight into decision-making on technology, few opportunities to gain

experience in policy-making and no stable research infrastructure regardless of the quality of the report (Smits et al., 1995, p. 292). A quality problem was that the scale of EU initiatives was too small (with a small budget) and varied, leading to insufficient quality, and therefore not taken seriously at the policy-making level.

In other words, there was a problem with the politicization of TA. In the US, OTA was criticized for being too close to democrats; in the EU, political dynamics made TA a residual instrument in the hands of parliaments. Advisory groups on ethics, sometimes falling under the label of “expert groups”³⁶ at the EC, are also vulnerable to being politicized. The politicization of ethical advice takes place when the norms defended for the legitimation of a particular ethical body (e.g. balance, independency or transparency) are eroded by partisan bias (Briggle, 2009, p. 314), although often presented as objective and neutral.

Depending on the composition balance of an ethics group, internal discussions could lead to almost opposite results depending on who was part of such groups (Evans, 2012, p. 95). In the EU, it was noted how ethics has been used as a way to push the economic and political visions of the EC on emerging technologies (Levidow & Carr, 1997, p. 38). In this rule of experts’ system, ethical decisions are claimed to be legitimated based on a “common morality”. In the case of the EU, this “common morality” consists of values enshrined in the EU Charter of Fundamental Rights and EU Treaties, representing European citizens. However, in fact, their legitimation is technocratic, as experts of these groups are not democratically elected, but selected based on their specialized knowledge (Evans, 2012, p. 129; Tallacchini 2009).³⁷ In this sense, ethics becomes an exercise of offering “technical” and “value-free” solutions to fixed facts, while addressing problems that were originally considered to involve conflicting values and political interests, such as in cases of Genetically Modified Organisms (GMOs) (Levidow & Carr, 1997; Wynne, 2001, 2005) or biotechnologies (Jasanoff, 2005; Tallacchini, 2015). Such values are then removed from consideration and only a pre-defined subset of them is taken for granted to achieve the most efficient solution.

To reflect on the ways technocracy and politicization are affecting AI ethics, it is necessary to first look at the responses that bioethics and TA, from the 1990s onward,

³² Examples are the more adversarial political culture and the stricter separation between the legislative and executive branches in the US compared to many European parliamentary systems, where the division of power is rather between government and opposition (Van Eijndhoven, 1997, p. 277; Vig 1992, p. 4).

³³ The OTA was much larger and “richer” than its European counterparts, allowing them to undertake many more projects per year.

³⁴ The research library that serves the US Congress. See (Kunkle, 1995, p. 175).

³⁵ See Sect. “Bioethics and the limits of principle-based ethics” and (Paula, 2008; Tallacchini, 2009, p. 287).

³⁶ Expert groups are consultative bodies set up by the EC to provide advice and expertise to the latter. EC expert groups are more or less the equivalent of advisory committees at a national level. See (Metz, 2015, p. 5).

³⁷ Technocracy here is opposed to representative democracy, according to which elected representatives are the main decision-makers in a given area regarding scientific or technological knowledge.

tried to implement to partly address the problem of expertise in the first place.

From ELSI and ELSA to RRI

Ethical, legal and social impacts research programs

It was the problem of expertise that mostly contributed to the crisis of traditional bioethics (Solomon, 2014) and the decline of traditional TA. These dynamics contributed to the emergence of different approaches, mostly motivated by the need to open up ethical discussions and involve more stakeholders, focusing on increased collaboration with experts from different fields (Zwart et al., 2014, p. 6).

In the US, many other (ethical) committees were created to support the work of the National Commissions,³⁸ including, later, the Ethical, Legal, and Social Implications (ELSI) Working Group (WG) of the Human Genome Project (HGP),³⁹ which became worldwide the basis for a distinctive research approach bearing the same name label. In contrast to expert-based traditional bioethics and TA (Hedgecoe, 2010), ELSI proposed an earlier upstream engagement, trying to “predict” or anticipate new technological developments also through the involvement of interdisciplinary stakeholders in the discussion. The idea was to collectively explore alternative paths to integrate societal concerns in the scientific practice, e.g. by allowing social scientists to engage with scientists in the laboratory.⁴⁰

The turn to ELSA in Europe (1994–2012) and the rise of RRI (2012–present)

ELSI initiatives were successful in proliferating outside the US, constituting a role model for ELSA in the EU and Member States’ research programs. Like the case of bioethics, ELSA in Europe played a different political role than ELSI in the US. ELSA was envisioned to contribute to integrating new scientific and technological development in European society, by legitimizing research and facilitating societal uptake of technological innovation, as a result of e.g. the EU backlash against agri-food technology in the 1990s (Levi-dow & Carr, 1997; Rodríguez et al., 2013).

³⁸ The National Commission of 1974 was succeeded by other comparable National Commissions, e.g. the “President’s Commission of the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research” (1980–1983), the “National Bioethics Advisory Commission” (1996–2001) or the “President’s Council of Bioethics (2001–2009).

³⁹ International project whose goal was to sequence the DNA genome.

⁴⁰ See CTA at Sect. “TA in the EU”.

ELS- aspects were considered for the first time in the EU in the context of the European Framework Programmes (FP) for Research and Technological Development, i.e. the EU’s main policy instrument to guide research and promote socio-technical integration and support R&D.⁴¹ After the spread of ELSA in the EU’s FPs, ELSA approaches started appearing in national funding bodies in several Member States since the 1990s, such as the Economic and Social Research Council Genomics Network in the UK or the Centre for Genomics and Society in the Netherlands.⁴²

Just like for TA, it is difficult to generalize about one single ELSA approach or object in Europe. The several initiatives that go under the name of ELSA share some general characteristics⁴³ but are also highly heterogeneous in terms of scope, expertise mobilized, and methods used to predict the future and engage stakeholders (Hilgartner et al., 2017, p. 828; Zwart & Nelis 2009, p. 541).

After a “golden decade” of ELSA approach (2002–2012), where different fields underwent a process of “elsification” (Zwart et al., 2014, p. 11), a new concept was introduced in the policy and research discourses of the EC, i.e. RRI,⁴⁴ presented as a groundbreaking change with the ELSA (recent) past (Owen, Pansera, Owen et al., 2021a, b; Owen, Schomberg, Owen et al., 2021a, b). In the period around 2011–2013, there was a growing impression at the EC that elsification of research was inadequate to address ethical aspects of emerging technologies, including AI and robotics. In the words of the main promoters of RRI,⁴⁵ a broader concept of responsibility and a more open design for innovation was needed (Owen, Schomberg, Owen et al., 2021a, b, p. 220).

RRI, compared to ELSA, offered a way not only to evaluate risks and ethical aspects of technologies, but to broaden the research process, promoting the active involvement of several stakeholders (e.g., from industry) that were absent in previous research programs of the EC. Such involvement

⁴¹ The Treaty on the Functioning of the European Union (TITLE XIX: Research and technological development and space, Article 179) states that “The Union shall have the objective of strengthening its scientific and technological bases by achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties”. Consolidated version of the Treaty on the Functioning of the European Union, *OJ C 326*, 26.10.2012, p. 47–390.

⁴² Some ELSA approaches, especially in the Netherlands, were particularly influenced by concepts and methods of TA.

⁴³ See Sect. “Ethical, Legal and Social Impacts research programs”.

⁴⁴ Other parallel initiatives, overlapping with RRI, emerged in this period, such as Responsible Innovation in the UK (Owen, Pansera, Owen et al., 2021a, b).

⁴⁵ Including, for instance, René Von Schomberg and Robert Madelin, then Director General at the EC’s DG Connect.

would help draw up different research agendas to better define and address “grand societal challenges” (Von Schomberg, 2012, p. 15). Additionally, instead of seeing ethical aspects as constraints or restrictions, RRI promoted a positive attitude to ethics, as an enabler of technology development (Von Schomberg, 2012, p. 16). Unlike ELSA, it targeted the whole innovation process and not just at the very end, from research to production and distribution, especially at early stages, in an anticipatory way (Von Schomberg, 2012).⁴⁶ In short, RRI sought to challenge, more ambitiously than ELSA, the seemingly linear and apolitical “technology/market dyad” that had dominated innovation in the previous decades, by re-configuring norms, institutions and political systems that govern innovation processes (Owen, Schomberg, et al., 2021, p. 221).

RRI was introduced to have a substantial role from the FP Horizon 2020 (H2020) onwards, as a top-down decision of the EC, and later established as an academic field and scientific-intellectual movement too.⁴⁷ Today, it is being discussed whether the initial promises and expectations around RRI are being kept (Owen, Schomberg, et al., 2021, p. 223).

ELSA-I, RRI and the limits of public participation

ELSI-A (as well as, later on, RRI in the EU) was developed to counter some of the problems of politicization and expertise of classical bioethics and TA highlighted in Sect. “TA and the politicization of ethical expertise”. While initially also ELSI, in the early 1990s, was meant to provide substantive ethical and social sciences expertise, its mission shifted a decade later, with a new emphasis on public participation, especially in the EU in the aftermath of the mad cow crisis and the rejection of GMOs (Hilgartner et al., 2017, p. 830). The alleged solution to this problem was to promote a two-way public engagement with science, abandoning the “public deficit” model explanation, which assumes that the general public lacked sufficient knowledge and understanding of basic science and technology (Flynn, 2007, p. 10). This lack of knowledge, in turn, would produce hostility towards science and technology, and therefore needs to be overcome to release the full potential benefits of the latter.

This abandonment, however, has been more apparent than real (Wynne, 2005). Some policy actors still misunderstand or ignore their causal role in public mistrust problems, not only because of individual responsibility, but also

“inadvertent reproduction of an established set of institutional reflexes and habits which inadvertently create public alienation” (Wynne, 2005, p. 217). Anchored to the old model of framing the public, they lack to acknowledge that the public is a construct, a reflection of the needs of those in power (Wynne, 2005, p. 218). It has been also noted how the effect of participatory efforts was limited, which was said to rarely produce significant benefits because of asymmetries of power (Hilgartner et al., 2017, p. 838), insufficiently deep and democratic participatory exercises (Wynne, 2006) or the risk of producing misleading representations of public preferences (Casiraghi et al., 2021; Hilgartner et al., 2017; Tait, 2009).

Despite the effort made to distinguish RRI from ELSA, so far, it has been pointed out how RRI has had similar problems as ELSA. RRI remains vaguely defined, or at least different visions are not necessarily consistent with each other.⁴⁸ One of the ways to define RRI is precisely the emphasis on public engagement, to associate citizens with the scientific process from the start, since integrating societal actors from different backgrounds is pictured as a way to nurture innovation and help to solve societal challenges. However, integrating societal actors with researchers, policymakers, and innovators is more complex than it seems in RRI discourses (Felt et al., 2016; Rommetveit et al., 2019, p. 89).

Similarly, the AI ethical agenda also (inadvertently?) further reflects problematic commitments. The deficit model comes under a new shape, from a deficit in scientific knowledge to a deficit in trust (Stilgoe & Guston, 2017, p. 862), as noted in the AI HLEG guidelines: “Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems – and the human beings behind them – being demonstrably worthy of trust, unwanted consequences may ensue and their uptake might be hindered, preventing the realisation of the potentially vast social and economic benefits that they can bring” (High-Level Expert Group on Artificial Intelligence, 2019a, pp. 4–5).

The main idea is that a lack of acceptance hinders innovation and blocks economic and competitive benefits, rather than signaling something else. The main concern is to smooth the integration of AI into society, rather than question its very necessity or impact.⁴⁹

⁴⁶ It was pointed out how some of these differences are rather caricatural and actually ELSA and RRI share many characteristics (Zwart et al., 2014, p. 13).

⁴⁷ See, for instance, the creation of dedicated journals (e.g. Journal of Responsible Innovation), handbooks or (online) training courses (e.g. <https://ocw.tudelft.nl/courses/responsible-innovation-building-tomorrows-responsible-firms/>). See also the idea of RRI as a scientific/intellectual movement in (Shanley, 2021).

⁴⁸ It was noted how RRI oscillates between an interactive process, responsible innovation or anticipatory governance (Rommetveit et al., 2019, p. 89).

⁴⁹ In biosciences, Huijter calls this approach “regulative ethics”, and, by extension, “regulative ethicists” are those that prepare and educate society to accept new technologies and instruct scientists how to behave responsibly, rather than challenging the hard core of science and its societal implications. Life scientists usually prefer to collaborate with this group rather than, say, STS scholars or healthcare

A further problem is that participatory exercises have been insufficiently deep and democratic (Wynne, 2006) or bear the risk of producing misleading representations of public preferences (Hilgartner et al., 2017, p. 830). The EU strategy on AI stresses the element of public participation as a salient feature of human-centric AI. As an example, in the AI HLEG guidelines, stakeholder participation is part of the requirement as ‘non-technical methods’ to ensure trustworthy AI (High-Level Expert Group on Artificial Intelligence, 2019a, p. 23). After the publication of the guidelines, the AI HLEG organized a piloting phase in which stakeholders were invited to provide feedback on the AI assessment list included in the guidelines, which resulted in the ALTAI assessment list (High-Level Expert Group on Artificial Intelligence, 2020).

Although these forms of participation should empower different groups and allow them to be heard in the debate on AI ethics, there are reasons to be skeptical about these outcomes (van Dijk et al., 2021, p. 11). First, concerning the piloting phase, only the last part of the document including guiding questions to operationalize the insights of the rest of the document was open to revision and comments, and not the whole document. The idea behind it was that “organizations working on AI” would “test” the assessment list of the ethics guidelines, which mostly included tech companies. Also, it was mentioned that the feedback loop would help to better understand how to implement the assessment list within an organization, thus being a vehicle to spread principles and requirements that were not open for discussion in the first place. By contrast, the first two parts of the guidelines, laying the conceptual foundations, requirements and principles, were *not* open to consultation. Second, this form of participation was limited to providing feedback, which could (or could not) be considered at the discretion of the EC.

These elements suggest how the ethics described rather empowers EU institutions while disempowering its citizens. Ethics documents often demarcate between “the public”, which needs to be passively surveyed and educated rather than genuinely involved, and stakeholders (e.g. scientists or businesspersons) who actively survey and educate (Greene et al., 2019, p. 2126) and are expert enough to provide valuable feedback. Public participation can be used by EU bodies, in these cases, rather for self-legitimation purposes and as a means of validation, or as a “smokescreen” to distort attention from the excessive involvement of the ICT industry in ethics initiatives (Metzinger, 2019; Wagner, 2018), resulting in too little room for, e.g., civil society to be part of these groups (Stolton, 2018).

ethicists that use unconventional bioethics approaches like ethics of care or narrative bioethics (Huijter, 2015, p. 94).

Conclusions

To wrap up, the article investigated some origins of the institutionalization of ethics of AI. This article does not aim to be exhaustive, but to highlight some patterns and characteristics that past examples of institutionalized approaches to technology assessment (some of which recall explicitly “ethics”, i.e. bioethics and ELSA/I, others not, i.e. TA and RRI) have in common with the ethics of AI in EU institutions. Bioethics, TA, ELSA-I and RRI have followed one another as institutionalized approaches to assess and govern the impacts of emerging technologies, for different reasons in the US as well as in the EU, especially in the life sciences. These approaches share similarities (including discourses and methods used, the importance given to expertise and/or to public participation) but they have also developed in opposition to their predecessors (e.g. ELSI vs. traditional bioethics, RRI vs. ELSA), to bridge their gaps and address their challenges.

First, the institutionalization of “traditional” bioethics was described, in the US and in the EU. While the origin of bioethics in the US was more related to controversies questioning the authority of scientists, and the role of the federal government in taking control of research ethics, the case of the EU was more related to solving the problem of democratic deficit and building a European identity. Second, the case of TA was presented. The history of TA in the US was inextricably linked with that of the OTA, and its role in mediating the tensions between the executive and legislative branches of the US federal government. In the EU, TA evolved differently from the initial conception of OTA, both for differences in political systems and contingent reasons of budget and capacity. Third, the developments of ELSI/A and RRI were described. ELSA in Europe played a different political role than ELSI in the US, supposed to “predict” or anticipate new technological developments also through the involvement of interdisciplinary stakeholders in the discussion. ELSA was envisioned to contribute to integrating new scientific and technological development in the European society, a role that was later taken up by RRI, that emphasized boosting innovation and addressing the “grand (socio-economic) challenges” of EU society through greater public participation and collaboration with the industry. Finally, the challenges faced by these forms of institutionalized ethics were related to the ones that AI ethics is facing today in the EU. The comparison showed that, despite changing labels and approaches, from bioethics to TA to ELSI/A and RRI, there are some recurring, unresolved challenges in ethics regarding methods employed, vulnerability to politicization and (misuse of) public participation.

From a methodological perspective, the value of principlism, and in general of the “quest” for the best principles

for AI ethics, should be questioned more. While the EU ethics claims to be fundamental rights-based and founded on EU values, it too often bluntly refers to bioethical principles, which are rooted in a cost-benefit perspective and risk-based discourses. While the very same idea of operationalizing principles may seem appealing from an engineering perspective, it poses serious challenges to the ways fundamental rights are articulated and operationalized in the EU context. One way to still draw on the knowledge of academic ethics could be to re-qualify the use of “moral philosophy”. Training in moral philosophy can still offer several contributions, such as framing debates and clarifying different positions, evaluating different positions and facilitating cooperation and engagement (Bietti, 2020).

Regarding the politicization of ethics, what is concerning is the exclusion, from ethical debates, of people and groups interested in discussing *ends* rather than how to develop ethical frameworks and principles to achieve, *ex-post*, predefined ones. Civil society groups are often marginalized in the debates, while they have been very active on the topic of e.g. ethics washing⁵⁰ and the text of the EC’s AI Act.⁵¹ Political and procedural questions are of crucial importance in institutionalized ethics (Briggle, 2009, p. 321). Despite the recent reforms of EC’s expert groups,⁵² primarily targeted at corporate dominance, there are still shortcomings in the way these groups function, especially regarding the balance of disciplinary backgrounds and political opinions of their members. In any case, deviation from rules to avoid disciplinary and bias imbalance should be enforced with sanctions.

As for public participation, it is too simplistic to simply call for more participation and involvement “of all stakeholders” or a “multi-stakeholder approach” “through the whole process of implementing AI systems” (High-Level Expert Group on Artificial Intelligence, 2019b, p. 37, 2019a, p. 19). There are different understandings, modalities and aims of participatory activities, from restoring trust in scientific institutions to legitimate policy commitments to encourage mutual understandings of different disciplines and interests (Felt et al., 2007, p. 56). Therefore, calls for more stakeholder involvement should be more specific and reflect on how participatory activities are performed, for whom and for what are they pursued, or who is supposed to participate and at which stage.

⁵⁰ For a plea for the use of human rights language rather than that of ethics, see <https://edri.org/our-work/attention-eu-regulators-we-need-more-than-ai-ethics-to-keep-us-safe/>.

⁵¹ The civil society statement on the AI Act is available at <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf>.

⁵² Expert groups are now subjected to horizontal rules established by Commission Decision C(2016)3301 of 30.5.2016 establishing horizontal rules on the creation and operation of Commission expert groups.

To conclude, there may be opportunities to change the governance of AI based on specific ethical considerations, but structural and discursive changes in institutionalized ethics are needed to avoid the pitfalls of the US and EU’s recent past. To implement these changes, two different strategies could be followed. On the one hand, if the language of ethics in the public discourse is considered useful, a different approach to this ethics, methodologically and institutionally, would be advisable. Conflicts and different standpoints about the governance of AI can be embraced rather than avoided by seeking consensus at any cost (Mouffe, 2001) in the name of “objective” values held by all EU citizens. On the other hand, a clarification of the use of the term “ethics” in general is needed,⁵³ since, having acquired so many different meanings in the documents and discourses analyzed, it has almost become an empty signifier (Mouffe, 2001; van Dijk et al., 2021) that can freely change its meaning (Buell, 2001). In some cases, this could mean a clarification of the philosophical basis (e.g., metaethical assumptions or normative theory) behind ethical claims that are made. In others, a “deflation” of the use of the term could be advisable. Human-centric and trustworthy approaches for AI and public participation initiatives, for example, do not always need to be named “ethical”, which can be misleading and lead to instrumentalization, but they could be robustly conceptualized by referring, for instance, to human rights, democracy and the rule of law.⁵⁴

Acknowledgements The author is grateful to Niels van Dijk, Paul De Hert and Mariachiara Tallacchini for their comments on an earlier draft of the article. This research was made possible through funding from the VUB’s Strategic Research Program ‘Articulating Law, Technology, Ethics, Politics: Issues of Enforcement and Jurisdiction of EU Data Protection Law under and beyond the General Data Protection Regulation (ALTEP-DP)’.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

⁵³ In the context of clinical ethics consultations, some authors have argued for a “terminological cleansing”, i.e. that the terms moral or ethics expertise should be replaced by more explicit and precise accounts of what the expertise involves. See (Ittis & Sheehan, 2016).

⁵⁴ This is the approach, for instance, taken by the Committee on Artificial Intelligence (CAI) of the Council of Europe. For a theoretical reflection in this vein, see (Nemitz, 2018).

References

- Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics* (8th ed.). Oxford University Press.
- Benatar, D. (2006). Bioethics and health and human rights: A critical view. *Journal of Medical Ethics*, 32(1), 17–20. <https://doi.org/10.1136/jme.2005.011775>.
- Bietti, E. (2020). From Ethics Washing to Ethics Bashing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*, 210–219.
- Bimber, B., & Guston, D. H. (1997). Introduction: The end of OTA and the future of technology assessment. *Technological Forecasting and Social Change*, 54(2–3), 125–130. [https://doi.org/10.1016/S0040-1625\(97\)81491-6](https://doi.org/10.1016/S0040-1625(97)81491-6).
- Breyer, H. (1992). *Committee on Energy, Research and Technology: Draft response to Bangemann report [CEC 1991]*
- Briggle, A. (2009). The kass council and the politicization of ethics advice. *Social Studies of Science*, 39(2), 309–326. <https://doi.org/10.1177/0306312708101048>.
- Buell, L. (2001). What we talk about when we talk about Ethics. In M. Garber, B. Hanssen, & R. L. Walkowitz (Eds.), *The turn to Ethics* (pp. 1–14). Routledge.
- Busby, H., Herve, T., & Mohr, A. (2008). Ethical EU law? The influence of the European group on ethics in science and new technologies. *European Law Review*, 2005, 803–842.
- Callahan, D. (1982). At the Center. *The Hastings Center Report*, 12(3), 4.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of Saint-Brieuc Bay. In J. Law (Ed.), *Power, action and belief: a new sociology of knowledge?* (pp. 196–223). Routledge. <https://doi.org/10.22394/0869-5377-2017-2-49-90>
- Casiraghi, S., Burgess, J. P., & Lidén, K. (2021). Social acceptance and border control technologies. In J. P. Burgess, & D. Kloza (Eds.), *Border Control and New Technologies* (pp. 99–115). Academic & Scientific Publishers. <https://doi.org/10.46944/9789461171375.6>.
- Chadwick, R., Gacesa, P., Brown, J., & Hedgecoe, A. (1998). *Sequencing the Human Genome: Scientific progress, Economic, ethical and social aspects*. European Parliament, Scientific and Technological Options Assessment.
- Chadwick, R., Hedgecoe, A., Isaksen, L., & Sarch, L. (1999). *Ethical issues in Research and Technology. Final study*. European Parliament, Scientific and Technological Options Assessment.
- Delacroix, S., & Wagner, B. (2021). Constructing a mutually supportive interface between ethics and regulation. *Computer Law and Security Review*, 40, 1–11. <https://doi.org/10.1016/j.clsr.2020.105520>.
- Eckenwiler, L. A., & Cohn, F. G. (Eds.). (2007). *The Ethics of Bioethics. Mapping the Moral Landscape*. The John Hopkins University Press.
- European Commission (1991). *Promoting the competitive environment for the industrial activities based on biotechnology within the community*.
- European Commission (2018a). *Artificial Intelligence for Europe*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>
- European Commission (2018b). *Coordinated Plan on Artificial Intelligence*. <https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence>
- European Commission (2020). *Ethics of Connected and Automated Vehicles. Recommendations on road safety, privacy, fairness, explainability and responsibility*. <https://doi.org/https://data.europa.eu/doi/10.2777/035239>
- European Data Protection Supervisor Ethics Advisory Group (2018). *Towards a digital ethics*. https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf
- European Group on Ethics in Science and New Technologies (2018). *Statement on Artificial Intelligence, Robotics and “Autonomous” Systems*. <https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF>
- European Parliament (1997). *European Parliament resolution on the mandate of the group of advisers on the ethical implications of biotechnology to the European Commission*.
- Evans, J. H. (2012). *The history and future of Bioethics. A sociological view*. Oxford University Press.
- Faden, R. R., & Beauchamp, T. L. (1986). *A history and theory of informed consent*. Oxford University Press.
- Felt, U., Wynne, B., Gonçalves, M. E., Jasanoff, S., Callon, M., Jepsen, M., Joly, P. B., Konopasek, Z., May, S., Neubauer, C., Rip, A., Siune, K., Stirling, A., & Tallacchini, M. (2007). *Taking European knowledge society seriously*. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/5d0e77c7-2948-4ef5-aec7-bd18efe3c442>
- Felt, U., Igelsböck, J., Schikowitz, A., & Völker, T. (2016). Transdisciplinary Sustainability Research in Practice: Between imaginaries of collective Experimentation and Entrenched Academic Value orders. *Science Technology and Human Values*, 41(4), 732–761.
- Floridi, L., COWLS, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schaffer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical Framework for a good AI society: Opportunities, Risks, Principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- Flynn, R. (2007). Risk and the public acceptance of new technologies. *Risk and the Public Acceptance of New Technologies*, 1–23. <https://doi.org/10.1057/9780230591288>.
- García del Blanco, I. (2020). *Report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL))*.
- Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2122–2131.
- Grunwald, A. (2009). Technology Assessment: Concepts and methods. *Philosophy of Technology and Engineering Sciences*, 9, 1103–1146. <https://doi.org/10.1016/B978-0-444-51667-1.50044-6>.
- Grunwald, A. (2019). The inherently democratic nature of technology assessment. *Science and Public Policy*, 46(5), 702–709. <https://doi.org/10.1093/scipol/scz023>.
- Hagendorff, T. (2020). The Ethics of AI Ethics: An evaluation of Guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>.
- Hedgecoe, A. (2010). Bioethics and the reinforcement of Socio-Technical Expectations. *Social Studies of Science*, 40(2), 163–186.
- High-Level Expert Group on Artificial Intelligence (2018). *Draft Ethics Guidelines for Trustworthy AI. Working Document for stakeholders’ consultation*.
- High-Level Expert Group on Artificial Intelligence (2019a). *Ethics Guidelines for Trustworthy AI*. <https://ec.europa.eu/futurium/en/ai-alliance-consultation>
- High-Level Expert Group on Artificial Intelligence (2019b). *Policy and Investment Recommendations for Trustworthy AI*.
- High-Level Expert Group on Artificial Intelligence (2020). *The Assessment List for Trustworthy Artificial Intelligence*. <https://data.europa.eu/doi/10.2759/002360>
- Hilgartner, S., Prainsack, B., & Hurlbut, J. B. (2017). Ethics as Governance in Genomics and Beyond. In U. Felt, R. Fouché, C. A.

- Miller, & L. Smith-Doerr (Eds.), *The handbook of Science and Technology Studies* (pp. 823–851). MIT Press.
- Huijter, M. (2015). Between Dreams and reality: The Dutch Approach to Genomics and Society. *Biosocieties*, 1(1), 91–95.
- Iltis, A. S., & Sheehan, M. (2016). Expertise, ethics expertise, and clinical ethics consultation: Achieving terminological clarity. *Journal of Medicine and Philosophy*, 41(4), 416–433.
- Jahr, F. (1927). Bio-Ethik: Eine Umschau über die ethischen Beziehungen des Menschen zu Tier und Pflanze. *Kosmos: Handweiser Für Naturfreunde*, 24(1), 2–4.
- Jasanoff, S. (2005). *Designs on Nature. Science and Democracy in Europe and the United States*. Princeton University Press.
- Jasanoff, S. (Ed.). (2011). *Reframing rights. Bioconstitutionalism in the genetic age*. The MIT Press.
- Jecker, N. S., Jonsen, A. R., & Pearlman, R. A. (1997). *Bioethics. An introduction to the history, methods and practice*. Jones and Bartlett Publishers.
- Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: The global landscape of ethics guidelines. *Nature Machine Intelligence*, 1, 389–399.
- Jonsen, A. R. (1998). *The birth of Bioethics*. Oxford University Press.
- Kunkle, G. C. (1995). New challenge or the past revisited? The Office of Technology Assessment in historical context. *Technology in Society*, 17(2), 175–196. [https://doi.org/10.1016/0160-791x\(95\)00007-e](https://doi.org/10.1016/0160-791x(95)00007-e).
- Levidow, L., & Carr, S. (1997). How biotechnology regulation sets a risk/ethics boundary. *Agriculture and Human Values*, 14(1), 29–43. <https://doi.org/10.1023/A:1007394812312>.
- McNeill, P. M. (1993). *The Ethics and Politics of Human Experimentation*. Cambridge University Press.
- Metz, J. (2015). *The European Commission, Expert Groups, and the policy process. Demystifying technocratic governance*. Palgrave Macmillan.
- Metzinger, T. (2019). Ethics Washing Made in Europe. *Der Tegesspiegel*.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>.
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI Ethics. *Minds and Machines*, 31(2), 239–256. <https://doi.org/10.1007/s11023-021-09563-w>.
- Mouffe, C. (2001). Which Ethics for Democracy? In M. Garber, B. Hanssen, & R. L. Walkowitz (Eds.), *The turn to Ethics* (pp. 85–94). Routledge.
- Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, 376(2133), 1–25. <https://doi.org/10.1098/rsta.2018.0089>.
- Owen, R., Pansera, M., Macnaghten, P., & Randles, S. (2021a). Organisational institutionalisation of responsible Innovation. *Research Policy*, 50(1), 1–13.
- Owen, R., Schomberg, R., Von, & Macnaghten, P. (2021b). An unfinished journey? Reflections on a decade of responsible research and innovation. *Journal of Responsible Innovation*, 8(2), 217–233. <https://doi.org/10.1080/23299460.2021b.1948789>.
- Paula, L. E. (2008). Ethics Committees, Public Debate and Regulation: An Evaluation of Policy Instruments in Bioethics Governance. In *Vrije Universiteit Amsterdam*.
- Plomer, A. (2008). The European group on ethics: Law, politics and the limits of moral integration in Europe. *European Law Journal*, 14(6), 839–859. <https://doi.org/10.1111/j.1468-0386.2008.00447.x>.
- Potter, V. R. (1970). Bioethics, the Science of Survival. *Perspectives in Biology and Medicine*, 14(1), 127–153.
- Reich, W. T. (1995). The Word “Bioethics”: The Struggle over its earliest meanings. *Kennedy Institute of Ethics Journal*, 5(1), 19–34.
- Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O’Sullivan, D., & Gordijn, B. (2018). Methods for Practising Ethics in Research and Innovation: A literature review, critical analysis and recommendations. *Science and Engineering Ethics*, 24(5), 1437–1481. <https://doi.org/10.1007/s11948-017-9961-8>.
- Rodríguez, H., Fisher, E., & Schuurbiens, D. (2013). Integrating science and society in European Framework Programmes: Trends in project-level solicitations. *Research Policy*, 42(5), 1126–1137. <https://doi.org/10.1016/j.respol.2013.02.006>.
- Rommetveit, K., Van Dijk, N., Gunnarsdóttir, K., O’riordan, K., Gutwirth, S., Strand, R., & Wynne, B. (2019). Working responsibly across boundaries? Some practical and theoretical lessons. In R. Von, Schomberg, & J. Hankins (Eds.), *International handbook on responsible Innovation: A Global Resource* (pp. 83–100). Edward Elgar.
- Rothman, D. J. (1991). *Strangers by the Bedside: A history of how Law and Bioethics Transformed Medical decision making*. Basic Books.
- Sartor, G. (2020). Artificial intelligence and human rights: Between law and ethics. *Maastricht Journal of European and Comparative Law*, 27(6), 705–719.
- Sass, H. M. (2007). Fritz Jahr’s 1927 Concept of Bioethics. *Kennedy Institute of Ethics Journal*, 17(4), 279–295.
- Schot, J., & Rip, A. (1997). The past and future of Constructive Technology Assessment. *Technological Forecasting and Social Change*, 54(2–3), 251–268. [https://doi.org/10.1016/S0040-1625\(96\)00180-1](https://doi.org/10.1016/S0040-1625(96)00180-1).
- Shanley, D. (2021). Imagining the future through revisiting the past: The value of history in thinking about R(R)I’s possible future(s). *Journal of Responsible Innovation*, 8(2), 234–253. <https://doi.org/10.1080/23299460.2021.1882748>.
- Smits, R., Leyten, J., & Den Hertog, P. (1995). Technology assessment and technology policy in Europe: New concepts, new goals, new infrastructures. *Policy Sciences*, 28(3), 271–299. <https://doi.org/10.1007/BF01000290>.
- Solomon, D. (2014). Bioethics and Culture: Understanding the Contemporary Crisis in Bioethics. *International Journal of Chinese & Comparative Philosophy of Medicine*, 12(2), 87–117. <https://doi.org/10.24112/ijccpm.121572>.
- Stilgoe, J., & Guston, D. H. (2017). Responsible Research and Innovation. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of Science and Technology Studies* (4th ed., pp. 853–880). MIT Press.
- Stolton, S. (2018). Artificial Intelligence presents’ black swan’ ethical issues, Commission report says. *Euractiv.Com*, 2–6.
- Tait, J. (2009). Upstream engagement and the governance of science. *EMBO Reports*, 10, 18–22.
- Tallacchini, M. (2009). Governing by values. EU ethics: Soft tool, hard effects. *Minerva*, 47(3), 281–306. <https://doi.org/10.1007/s11024-009-9127-1>.
- Tallacchini, M. (2015). To bind or not to bind? European Ethics as Soft Law. In S. Hilgartner, C. A. Miller, & R. Hagendijk (Eds.), *Science and Democracy. Making knowledge and making power in the biosciences and beyond* (pp. 156–175). Routledge.
- Turner, L. (2009). Does bioethics exist? *Journal of Medical Ethics*, 35(12), 778–780. <https://doi.org/10.1136/jme.2008.028605>.
- van Dijk, N., Casiraghi, S., & Gutwirth, S. (2021). The ‘Ethification’ of ICT Governance. Artificial Intelligence and Data Protection in the European Union. *Computer Law & Security Review*, 43, 105597. <https://doi.org/10.1016/j.clsr.2021.105597>.
- 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).
- Vig, N. J. (1992). Parliamentary technology assessment in Europe: Comparative evolution. *Impact Assessment*, 10(4), 3–24. <https://doi.org/10.1080/07349165.1992.9725818>.

- Von Schomberg, R. (2012). René von Schomberg on responsible innovation. See ethics as a stimulus and not as an obstacle. *LEV*, 14–18.
- Wagner, B. (2018). Ethics as an escape from Regulation: From ethics-washing to ethics-shopping? In M. Hildebrandt (Ed.), *Being profiled. Cogitas Ergo Sum* (pp. 84–89). Amsterdam University Press.
- Wynne, B. (2001). Creating Public Alienation: Expert cultures of risk and Ethics on GMOs. *Science as Culture*, 10(4), 445–481. <https://doi.org/10.1080/09505430120093586>.
- Wynne, B. (2005). Public Engagement as a Means of restoring Public Trust in Science - Hitting the Notes, but missing the music? *Community Genetics*, 9(3), 211–220.
- Wynne, B. (2006). Public Engagement as Means of restoring Trust in Science? Hitting the Notes, but missing the music. *Community Genetics*, 3(9), 211–220.
- Yesley, M. S. (2008). What's ELSI got to do with it? Bioethics and the Human Genome Project. *New Genetics and Society*, 27(1), 1–6. <https://doi.org/10.1080/14636770701843527>.
- Yeung, K., Howes, A., & Pogrebna, G. (2020). Why industry self-regulation will not deliver “Ethical AI”: A call for legally mandated techniques of “Human Rights by Design. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI*. Oxford University Press.
- Zwart, H., & Nelis, A. (2009). What is ELSA Genomics? *EMBO Reports*, 10(6), 540–544.
- Zwart, H., Landeweerd, L., & van Rooij, A. (2014). Adapt or perish? Assessing the recent shift in the european research funding arena from “ELSA” to RRI. *Life Sciences Society and Policy*, 10(11), 1–19.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.