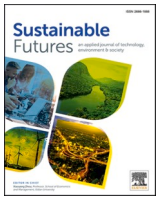
Sustainable Futures 4 (2022) 100086

Contents lists available at ScienceDirect

Sustainable Futures

journal homepage: www.sciencedirect.com/journal/sustainable-futures



Towards a unified list of ethical principles for emerging technologies. An analysis of four European reports on molecular biotechnology and artificial intelligence

Joachim Boldt [a,#], Elisa Orrù [b,#,*]

[a] Department of Medical Ethics and the History of Medicine, CIBSS – Centre for Integrative Biological Signalling Studies, University of Freiburg, 79104, Freiburg, Germany
[b] Centre for Security and Society and Philosophy Department, University of Freiburg, 79098, Freiburg, Germany

ARTICLE INFO

Keywords:
Technology governance
Technology assessment framework
Artificial intelligence ethics
Bioethics
Emerging technologies

ABSTRACT

Artificial intelligence (AI) and molecular biotechnologies (MB) are among the most promising, but also ethically hotly debated emerging technologies. In both fields, several ethics reports, which invoke lists of ethics principles, have been put forward. These reports and the principles lists are technology specific. This article aims to contribute to the ongoing debate on ethics of emerging technologies by comparatively analysing four European ethics reports from the two technology fields. Adopting a qualitative and in-depth approach, the article highlights how ethics principles from MB can inform AI ethics and vice versa. By synthesizing the respective ethical cores of the principles included in the analysed reports, the article derives, moreover, a unified list of principles for assessing emerging technologies. The suggested list consists of nine principles: autonomy; individual and social well-being and prevention of harm; reliability, safety and security; informational privacy; transparency; accountability; communication, participation and democracy; justice, fairness, and non-discrimination; sustainability.

1. Introduction

1.1. Ethics principles and technology governance in culturally heterogeneous democratic societies

Assessing and developing emerging technologies in accordance with ethical values and standards helps to foster their beneficial impact on society, avoid harm, and warrant public trust. Accordingly, a variety of approaches and tools have been devised on how to assess the societal and ethical impact of emerging technologies. Examples include the concept of "Responsible Research and Innovation" (RRI) [1–3], Technology Assessment (TA) schemes [4,5], and Ethical Foresight Analysis (EFA) [6,7]. In line with these efforts, the EU has developed a "Science with and for Society" programme as part of its research and innovation funding strategy [8], and routinely subjects research applications to an ethics appraisal procedure [9,10].

Often, lists of ethical principles are used within these and other broader and governance oriented approaches. These approaches are "agnostic", though, with regard to what list and which principles are included. The RRI, TA and EFA frameworks as such, for instance, can be built independently of the content of specific ethics principles lists and can work with different kinds of such lists [5,7]. Conversely, lists of principles can be set up independently of RRI, TA and EFA frameworks and can be used within different frameworks of this kind.

It has been argued that applying lists of ethics principles as part of ethics assessment procedures suffers from a number of drawbacks. It has been maintained, for example, that principles may come into conflict with each other, that these lists of principles lack theoretical unity and that they may reinforce a questionable box-ticking attitude towards ethical challenges [11–13].

Nonetheless, lists of principles are valuable and necessary whenever ethics assessments are commissioned as part of a political or regulatory process in socially and culturally heterogeneous democratic societies [14]. They should not be regarded as definite solutions to ethical debate

* Corresponding author.
E-mail address: elisa.orru@css.uni-freiburg.de (E. Orrù).
# The authors wish it to be known that, in their opinion, both authors should be regarded as joint First Authors having equally contributed to the article.

https://doi.org/10.1016/j.sftr.2022.100086
Received 16 May 2022; Received in revised form 4 June 2022; Accepted 5 June 2022
Available online 6 June 2022

but rather as tools within deliberative democratic processes, provided by, e.g., interprofessional and interdisciplinary ethics committees, to enable ethically informed decisions at a specific time and place.

Different ethics theories recommend different main principles,[i] different cultures and countries have different traditions of societally embedded ethical points of view, and different legal and governance frameworks add further, and differing, constraints. Lists of principles can help to identify, systematize and sample values and ethical convictions that may be found, in different guises and hierarchical order, in many of the more abstract ethics theories, and in different ethical traditions and cultures. Principles lists thus enable cross-cultural dialogue and help to pinpoint issues that can be agreed upon and others, where disagreement pertains. They thus provide a starting point for further democratic decision-making. What is more, principles lists can help to identify at what points and in what ways a technological application fails to live up to agreed-upon ethical standards. Principles lists thus can help to further refine or change applications of a technology to render them ethically more acceptable, or to ameliorate ethically unfavourable consequences of an application.

1.2. Ethics of artificial intelligence versus ethics of molecular biotechnologies

Artificial intelligence (AI) and molecular biotechnologies (MB) are two of the most promising and, at the same time, controversial emerging technologies today. Both these technologies are among the ten most important technologies identified by the OECD in its Science, Technology and Innovation Outlook 2016 [15].[ii] Biotechnology and Information Technology are part of the "NBIC" technologies, a popular acronym to denote nano-, bio-, information-, and cognitive technologies and to highlight their potential to significantly shape the future of our societies [16,17]. Both of these fields of technology do have a long pedigree of research and technological advances, such as 20th century genetic engineering and IT. Nonetheless, recent advances in, for example, gene editing and machine learning enable developing new ranges of applications that have hitherto been impossible to devise. In this sense, albeit having a long history, MB and AI can qualify as "emerging" technologies.

At first sight, these two technologies appear to be very different. While AI is concerned with information, algorithms, and artefacts, MB involves living beings, growth, and evolutionary processes. One may thus suspect different ethics issues and principles to be relevant for and characteristic of each field. Nonetheless, applications within each field can differ significantly. AI can be part of face recognition for social engineering as well as robotics in automated manufacturing plants. MB applications range from, for example, releasing genetically engineered mosquitos into the environment to diagnostic or therapeutic use of bacterial DNA compartments in the human body. In addition, AI applications and approaches can be used within MB and vice versa. AI machine-learning algorithms are used to analyse large amounts of data regarding the movement and interaction of intracellular molecules in "omics" research. Engineering methods and principles such as design, orthogonality, standardisation, and modularisation are imported from traditional engineering disciplines, including AI, to the realm of MB, and young and enthusiastic practitioners transfer IT hacker ideals to genetic engineering [18,19]. With regard to future applications, van Est et al. predict "info-tech interventions" into animal and human organisms, such as smart e-pills and brain-machine-interfaces [16].

This combination of internal variety of applications within each field and convergences and overlaps of applications and approaches between those fields makes a comparison between ethics principles of the two fields of technology a promising endeavour. It is our working hypothesis that existing differences give rise to partly heterogeneous principles in the two technology fields, while similarities and synergisms make these different principles relevant for the other area as well. We expect that sets of principles established in one area do not completely overlap with sets of principles developed in the other area. A comparison between those sets can highlight principles currently underrepresented and yet relevant in the respective technology fields.

Both within the ethics of MB and AI, the principles approach in medical ethics has been taken as an inspiration for devising lists of technology specific ethics principles. The medical ethics approach posits autonomy, beneficence, non-maleficence, and justice as the four principles needed to cover ethical issues in the clinic and healthcare more generally [20]. In the debate on ethics of AI, Floridi et al. [21] amended these four canonical principles with a fifth principle, namely "explicability".[iii] Regarding MB, the U.S. Presidential Commission for the Study of Bioethical Issues proposed a set of principles that adds responsible stewardship and democratic deliberation to, roughly speaking, a subset of autonomy, a wide understanding of benefit, and justice [22,23]. Other reports and publications devise own principle catalogues, often substantially diverging from the medical ethics list. The reports we chose for analysis are cases in point. Nonetheless, the elaborate debate on principles in medical ethics can still be relevant for devising principles for emerging technologies. We thus refer to the medical ethics debate in a few cases below (Section 3).

1.3. Aims of the paper

On this background, our contribution to the debate is twofold. Firstly, we aim to illustrate that a synthesis of existing ethical principles for AI and MB respectively allows for appreciating how the ethics debates in both these fields can profit from each other and to overcome current blind spots in the ethics of AI and MB respectively. To the best of our knowledge, the present paper is the first one to offer a comparison of ethical reports and principles in the two domains and to highlight which lessons each of the two areas of ethical debate can learn from the other. We maintain, secondly, that by drawing on ethics principles proposed for evaluating AI and MB, it becomes possible to compile a list of ethics principles capable of covering further emerging technologies and their applications.

We do not claim, though, that each principle will necessarily be relevant for every new emerging technology and every application of an emerging technology. For example, while the agent-like features of some AI applications raise issues of accountability, MB technologies usually do not. The same problem of relevance can arise for different applications within one and the same field of technology. In some of those cases, though, a principle initially thought to be irrelevant may, on closer inspection, turn out to be relevant in a different, distinct manner. For example, accountability is in fact an important part of devising ethically acceptable implementation and governance processes concerning MB.

Thus, even if a principle initially does not appear to be relevant for an application or technology field, having included it in a principles list ensures checking its relevance. A comprehensive list of principles guarantees this checking procedure, not excluding, of course, the possible result that with regard to a specific application or technology, a principle finally turns out not to be relevant.

[i] In addition to the traditional ethical theories expected to inform responsible research and innovation, namely deontology, utilitarianism and virtue ethics, more recent approaches such as decolonial, feminist, informational and deliberative approaches have been suggested [51,52,53,54].

[ii] The OECD uses the term „synthetic biology" instead of „molecular biotechnology". „Synthetic biology" is sometimes defined narrowly, restricted to modularization and standardization approaches. In order to avoid misunderstandings caused by this narrow definition of the term, we adopt the more neutral and broad term „molecular biotechnology"

[iii] The challenges and shortcomings of a translation of ethical principles from the biomedical to the AI domain are discussed by Mittelstadt [55].

The article is structured as follows: In Section 2, we introduce the four selected reports and identify ethics principles put forward by these reports. In Section 3, we systematize the emerging principles into coherent clusters, explore the possible contribution of principles that are only introduced in one of the fields to the other technology area, and suggest a consolidated list of principles and their specifications. We conclude this section with additional remarks on conceptual framings and their relationship to ethical principles. In Section 4, we summarise our findings, discuss limitations and identify areas where further debate is needed.

2. Presentation of reports

2.1. Choosing relevant reports

The landscape of ethics reports for MB, on the one hand, and AI, on the other hand, has developed differently. Regarding MB, early reports and guidelines emerge with the advent of genetic engineering in the last decades of the 20th century. These reports usually focus on specific areas of application and specific techniques such as plant genetic engineering or cloning. A number of reports that take a more comprehensive perspective on MB appeared between 2005 and 2015, when researchers coined the term "synthetic biology" to promote an approach to "editing" and "writing" gene sequences and genomes inspired by engineering principles and methods such as design, standardization and modularization [24]. These are the reports relevant for our purposes. Later reports again tend to focus on specific areas of application and specific techniques such as safety aspects of gain of function research and CRISPR/Cas genome editing.

By contrast, the corpus of ethical reports regarding AI first started to establish and develop around 2010 and has expanded at a rapid pace since 2016 [25]. Currently, more than 160 ethical guidelines for AI exist globally, provided by such different actors as private organisations, civil society groups, government agencies, academia, professional associations etc. [26].

In order to analyse ethics principles proposed in these fields, we opted for a qualitative and in-depth approach consisting of a detailed comparison of four selected reports. This approach allows for appreciating differences in the way principles are understood, inconsistencies of terminology across reports, and reinterpretation and rearrangement of principles based on the explanations and perspectives given by the reports themselves. It thus allows us to arrive at a better understanding and explanation of the resulting principles list. As a drawback, the results of this approach cannot claim to be a representative depiction of *all* relevant principles proposed in the debates. There may be principles in other European reports that remained undetected, and there almost certainly will be other principles in reports worldwide. However, given our specific aims, we regard the advantages of close-reading a small number of reports to outweigh the drawbacks.

In a similar vein, we have not sampled and chosen reports following an all-encompassing database research and a weighing of reports according to quantitative measures such as citation counts. Instead, we have carried out web searches and, complementarily, made use of our respective scholarly expertise in the domains of ethics of AI and MB. In addition, we have validated our sample of AI reports by comparing it to the overview of existing AI ethical guidelines provided by Jobin et. al [25], integrated with a search on the *AI Ethics Guidelines Global Inventory* run by Algorithm Watch [26].[iv] No such overview exists with regard to MB ethics principles. The resulting set of reports was weighed according to the following criteria:

Firstly, we prioritized reports which take account of the current academic ethical debate and are directed at informing politics and the public. This implied privileging recent reports supported by public funds and produced by (groups of) experts including ethics scholars. By contrast, we attached less weight to reports that are either produced by or target the private sector, or have a purely academic character, or adopt an approach in which ethical considerations play an ancillary or derivative role.

Secondly, we favoured reports with a policy focus beyond national debates and national regulations, but restricted to a European context. More specifically, we opted to grant special (yet not exclusive) consideration to reports commissioned by the EC. This is why otherwise interesting and important reports, such as the UNESCO *Recommendation on the Ethics of Artificial Intelligence* [27], adopted in November 2021, with its explicit global scope, were prioritized low. The already existing European Commission ethics-related funding schemes, research-related ethics obligations and research proposal assessments, including EC commissioned reports, provide an existing framework of international and cross-cultural governance and ethics cooperation. In the field of AI, a genuinely intercultural and promising dialogue around the cross-cultural stand of ethics principles has already started [28,29]. This European framework is, on the one hand, a valuable source for relevant ethics reflection and principles, and, on the other hand, contains institutions and tools that may make use of or further refine unified principles lists of the sort we aim to contribute to.

Thirdly, we gave preference to reports that are sufficiently broad in scope both from a technological and from an ethical point of view. From a technological point of view, we selected guidelines targeting a wide range of techniques and applications of the respective technology, and thus discarded reports which focussed only on, for example, CRISPR/Cas, plant genetic engineering, robotics, or autonomous driving. Regarding the ethical scope, we excluded reports exclusively based on one or a few ethical principles, such as non-discrimination or privacy.

2.2. Two reports on ethics of MB

2.2.1. Introduction

Ethical concerns, guidelines, and debates accompany genetic engineering from its very beginnings. When it became possible to transfer DNA from different donor organisms into host cells via recombinant DNA techniques in the 1970s, this imminently caused safety concerns and led to the now famous conference on recombinant DNA at Asilomar. In the early years of the 21st century, the debate intensified again with the advent of synthetic biology, a genetic engineering approach aiming at in-depth rational design of genomes via gene synthesis, and thanks to CRISPR/Cas and other gene editing techniques it has not come to a halt since.

The term "Molecular Biotechnologies" (MB) covers all these genome-directed editing technologies and its applications. Applications range from energy, to environment, to food and agriculture. They also include human health applications that make use of altered non-human DNA, genomes and organisms for human health purposes, as well as direct interventions into the human genome such as gene therapy, human genetic germline interventions, and human genetic enhancement [30].[v] Poster child applications of MB include the re-engineering of the yeast genome and its metabolism to enable the production of artemisinin, a substance used to treat malaria, and devising synthetic bacterial DNA compartments to sense disease within the human body [31–33].

[iv] Other existing overviews of ethical guidelines for AI include: [56], who also provide an overview of existing meta-studies on AI ethics reports, [21,57,58]. All of them, by contrast to our approach, have privileged breadth of scope over depth of analysis.

[v] It should be noted that broad reports on MB often include direct interventions into the human genome in their definition of „MB", but usually do not focus on these applications. Gene therapy, genetic interventions into the human germline, and human genetic enhancement each have sparked an ethical debate of its own, to which MB reports cannot and do not attempt to fully live up to.

Two reports were selected for closer scrutiny. The first one is the Opinion Paper 25 of the European Group on Ethics in Science and New Technologies (EGE) on synthetic biology [34]. The EGE is an ethics advisory body to the European Commission. The second report, "Emerging biotechnologies: technology, choice and the public good", was published by the British Nuffield Council on Bioethics (NCoB), a UK institution comparable to other countries' national ethics councils [35].

2.2.2. Ethics of synthetic biology (EGE)

The EGE's opinion paper 25 emphasizes the need for a unified and consistent ethical frame for evaluating synthetic biology. It chooses the concept of **human dignity** as a rather philosophical starting point of its assessment. The EGE claims that other, lower level ethical principles can be derived from human dignity, among them protective principles regarding animals and the environment. Quoting Cheshire [36], the EGE describes human dignity as "the exalted moral status which every being of human origin uniquely possesses. Human dignity is a given reality, intrinsic to human substance, and not contingent upon any functional capacities, which vary in degree. (…) The possession of human dignity carries certain immutable moral obligations. These include, concerning the treatment of all other human beings, the duty to preserve life, liberty, and the security of persons, and concerning animals and nature, responsibilities of stewardship." [34] (39).

As a next step, the EGE names a number of ethical issues and concerns that are supposed to be relevant for evaluating synthetic biology. The EGE introduces transparency and societal participation regarding research and development, application-specific ethics issues of potential concern, and the potentially ethically worrisome way in which synthetic biology makes use of the terms "life" and "nature" [34] (39-41). In the following, though, the EGE does not make systematic use of these initial considerations. As a matter of fact, the structure of the main part of the evaluation is guided by these aspects and issues: biosafety, biosecurity, governance, intellectual property, trade and global justice, science communication (dialogue between science and society), and research funding [34] (42-55).

In the course of its argument, the EGE refers to the ethico-philosophical debate between **anthropocentric and eco-centric positions** in ethics and argues that even an eco-centric approach need not lead to radical opposition to synthetic biology, but can allow for trade-offs [34] (41f). The section on biosafety also introduces and explains the **precautionary principle** [34] (42f.). It is emphasized that this principle does not simply favor the status quo, as passivity can also involve risks. However, this principle leads the EGE to demand, among other things, long-term studies on the impact of synthetic organisms on the environment before such organisms are approved for uncontained use.

Besides its affinity for ethico-philosophical issues of synthetic biology and a legal focus, which comes to the fore when the EGE discusses intellectual property, the EGE is thus close to nature-conserving and technology-critical positions, but at the same time strives to demonstrate the general compatibility of these positions with technological progress.

2.2.3. Emerging biotechnologies (NCoB)

The NCoBl's report differs from the statements of the EGE first of all by a social science background. While the EGE has philosophy and law as its points of reference, the NCoB understands newly emerging biotechnologies primarily as social and societal phenomena that need to be examined with the means of the social sciences. Like the EGE, the NCoB emphasizes that predictions about the development and the benefits and harms of new biotechnologies are subject to **uncertainty**, especially regarding future development, ambiguity of attached values, and transformative potential [35] (41-43).

The NCoB generally emphasizes that besides research and technology and its impact on human beings and the environment, the way in which a new technology is **conceptualized** ("framed", "represented"), the tradition in which new biotechnologies are placed, and the fields of application that are highlighted also need to be examined. This parallels and widens the EGE's interest in the way synthetic biology makes use of the terms "life" and "nature".

This demand is linked to the NCoB's assumption that there is no single correct way to understand a technology and its value. To take up an example the NCoB invokes itself, a technology that promises to prolong human life may appear beneficial to one person, but another person may see it as a harmful technology that leads to overpopulation and injustice [35] (45). According to the NCoB, both "values" (perhaps the term "view" would be more appropriate here) should be considered when assessing the (respective) technology. The NCoB stresses that certain cultures and systems of faith may ascribe special moral status to living beings and that technologies are subject to conceptual framings regarding the meaning of the technology itself, its tradition, and its fields of envisaged applications. In addition, the NCoB assumes that a new technology is seldom without competition when it comes to the question of how a societal challenge can be met. There are usually alternative technologies or social measures that could also help solve these problems [35] (9, 14f).

These **diversities of values** and alternative technologies and other **alternative means** to tackle societal challenges all need to be mapped and weighed up, before ethically justified decisions can be made, the NCoB claims. This is best done via a "public discourse ethics" oriented at the "public good" [35] (61f., 67f).

From these initial considerations, the NCoB derives a list of ethical principles that is largely procedurally oriented. These procedural principles ("procedural virtues", as the Council has it) are supplemented by three rather material principles. These latter three principles are **equity, solidarity** and **sustainability**. Equity entails equal respect for interests and preferences of others and their views on distributive justice, solidarity requires avoiding social divisiveness and active promotion of the welfare of those who less advantaged, sustainability is meant to cover sustainable use of natural resources [35] (63f).

The procedural principles are introduced as a discourse-ethical "application" of the three material principles. The procedural principles are: openness and inclusion, accountability, public reasoning, candor, enablement, and caution. Openness and inclusion require that all relevant stakeholders and the public should have access to information on new biotechnology and be able to participate in governance. Accountability demands that the public and social actors acknowledge and accept, where responsibility for governance is located, and that they are able to revise the allocation of responsibility via democratic procedures. Public reasoning encompasses impartial reasoning on the value and impact of a biotechnology. The reasoning should be oriented towards common ground rather than stakeholder specific interests. Candor asks for truthful representation especially of uncertainties relating to the impact of new biotechnologies. Enablement entails enabling wider political debate on new biotechnologies, including debating alternative social and technological options. Caution demands that if the impact of a new biotechnology is uncertain and the evaluation is ambiguous, deep and extensive knowledge should be gathered prior to making governance decisions [35] (68-71).

2.3. Two reports on AI ethics

2.3.1. Introduction

Although research on AI has been in place for 70 years now, only the recent exponential increase in data storage and computing capacities has enabled this technology area to exploit its potential and significantly expand its area of application. At a rapid pace, AI systems have intruded many areas of human activities, including transport, industry, health care, learning and teaching, the arts, finance and insurance, the military and law enforcement. To mention only applications in very mundane and everyday activities, AI technologies today support internet searches, navigation systems, voice assistants and photo tagging on social media [37] (3–4).

Today, "AI" is used to refer to software and hardware systems that perform activities and cognitive tasks traditionally considered to be exclusively human. Given a goal, these systems can autonomously acquire and interpret information, formulate suggestions or take decisions based on the data processed. To perform their tasks, contemporary AI systems commonly rely on techniques such as machine learning (including deep learning and reinforcement learning), machine reasoning and robotics [38] (6).

This session presents two documents that focus on ethical guidelines for AI technologies. The first one, the *Ethics Guidelines for Trustworthy AI*, was made public in April 2019. The guidelines have been produced by a group of 52 experts appointed by the EC in June 2018, the High-Level Expert Group on AI (AI HLEG). Notwithstanding the criticisms it has been subjected to, the AI HLEG report is widely acknowledged as one of the most impactful and advanced ethical tools for the evaluation of AI systems. Even critics concede that the guidelines are the currently best available basis for discussing ethical issues related to AI [39]. Given the anchoring of the guidelines into the EU institutional framework, they can be reasonably assumed to have a significant impact on EU research policy in the years to come, and to become (one of) the most important benchmark for the assessment of EU-funded research in the field of AI.

The second document, *Understanding artificial intelligence ethics and safety. A guide for the responsible design and implementation of AI systems in the public sector*, was published by the Alan Turing Institute (ATI), the UK national institute for data science and artificial intelligence, in 2019. For the effort endorsed to build a comprehensive framework for ethical AI artefacts, this report is comparable to the AI HLEG guidelines. Moreover, although the mentioned principles are suitable for AI applications in general, this document has a specific focus on AI uses for the public sector, which are among the ethically most challenging applications of AI.[vi]

2.3.2. AI HLEG guidelines for trustworthy AI

The *Guidelines for Trustworthy AI* present a set of four ethical principles that should govern the development and use of AI technologies. On the one hand, these four principles are grounded in existing human rights law, while, on the other hand, they inform a catalogue of seven key requirements for the realization of trustworthy AI artefacts. The specification into requirements shall enable the translation of the ethical principles into practical recommendations for assessment and evaluation of AI technologies. Thus, the guidelines build a three-level structure, in which ethical principles follow from human rights and the seven key requirements follow from ethical principles (see Table 1 below).

Table 1
Dependencies between principles and requirements as described by the AI HLEG guidelines.

AI HLEG ethical principles	AI HLEG key requirements
Respect for human autonomy	– Human agency and oversight
Prevention of harm	– Robustness – Privacy and data governance – Societal and environmental well-being (see also below)
Fairness	– Diversity and non-discrimination – Societal and environmental well-being (see also above) – Accountability
Explicability	– Transparency

By adopting human rights as a foundation for the guidelines, the HLEG explicitly opts for a "human-centric" perspective [40] (4). In a way that parallels the EGE report, the AI HLEG grounds its ethics guidelines in human dignity, namely the conviction that humans enjoy a unique moral status and that the respect for this moral status shall be ensured in all aspects of life. Echoing the Kantian categorical imperative, human dignity is specified by the AI HLEG as the requirement to treat human beings as subjects, rather than as objects (the Kantian terminology would be to treat persons as "ends" versus "pure means"). Although human dignity is attached to each human being as an individual moral subject, the HLEG does not draw upon it as an individualistic framework. By contrast, the guidelines understand this specific anthropocentric approach as a way to ensure that AI systems shall be in the service of "humanity and the common good" [40] (4) and consider individuals in their economic, societal and political context. This human dignity related approach inspires the four ethical principles that shall govern AI uses, namely: respect for human autonomy, prevention of harm, fairness and explicability [39] (11-13).

At the third level, the guidelines transform the mentioned ethical principles into key requirements for trustworthy (i.e. ethical and robust) AI systems. The rationale for this further step is to specify the ethical principles in order to make them operable in practice. From the ethical principle of respect for human autonomy, thus, the AI HLEG derives the requirement of human agency and oversight on AI artefacts. According to this requirement, AI should not substitute but instead support human decision making, by fostering users' agency, respect for human rights and for a democratic and equitable society. AI systems, moreover, shall allow for human oversight, as a way to check and ensure that human agency is actually respected by them [40] (15-16). The principle of prevention of harm founds the requirements of technical robustness and safety, privacy and data governance and societal and environmental well-being (which is in part also derived from the principle of fairness). Technical robustness and safety imply a proactive approach to risks related to the use of AI systems aimed at minimising intended and unintended harms, as well as ensuring that the system works as expected and delivers accurate results [40] (16-17). Privacy and data governance require that personal data are adequately protected, but also that the system uses accurate and non biased data and that it offers secure data access and management procedures [40] (17). Societal and environmental well-being prescribes to also take into account environmental protection, the interests of broader society and of future generations [40] (19). The ethical principle of fairness implies diversity, non-discrimination and fairness, as well as accountability. The requirement of diversity, non-discrimination and fairness is not limited to avoidance of biases, but also includes the involvement of relevant stakeholders during the development process of AI systems and making AI systems accessible by everybody through inclusive design [40] (18-19). Accountability aims to ensure that AI systems and their outcome be auditable, that trade-offs and negative impacts are made transparent and that effective redress mechanisms in case of adverse impact of the systems are in place [40] (19-20). Finally, the ethical principle of explicability is based on the requirement of transparency. This final requirement prescribes that the data collection and processing activities are documented in order to allow traceability, that both processes and outcomes are explainable and that communication while interacting with humans is transparent, both about the non-human nature of the system and about its limitations [40] (18).

2.3.2. The ATI report

Interestingly, the report by the ATI identifies as sources of AI ethics two strains: Bioethics and the human rights discourse. From the former, AI ethics draws the focus on safeguarding individuals' interests and well-being. The latter stresses the importance of human dignity, understood as a concept including freedom, autonomy, human flourishing and self-determination [41] (8-9).

The need for AI ethics derives, according to the report, from the

[vi] Such uses (including AI-based tracking and identification techniques, as well as citizens' scoring), although not specifically labelled as "public uses" are mentioned by the AI HLEG guidelines as applications raising particular concerns [40] (334).

broad range of potential harms that can arise from AI technologies, especially from their deployment in the public sector. These harms include, among others, bias and discrimination, the denial of individual autonomy and rights, invasion of privacy and disintegration of social connection, as well as unreliable outcomes [41] (4-5).

The ATI report, too, builds a multi-level ethical framework. It distinguishes between three levels. First, it identifies an ethical framework to inspire the work, with which developers of AI technologies should become familiar before starting to work on a project. Second, it considers a set of "actionable principles" to ensure that the AI project being developed is ethically permissible. Finally, it describes a PBG-Framework (Process-Based-Governance Framework), which identifies procedural tools to implement the two bulks of principles above through transparent and justifiable workflow processes [41] (7).

Although the first level is said to consist of "values", no ethical principles in the common sense of the expressions are formulated there. Instead, this level describes the general attitude and normative background which is favourable for developing ethical AI systems. More to the point for our discussion is the second level, which includes the so-called FAST Track principles: **fairness, accountability, sustainability** and **transparency** [41] (11 ff.).

The main function of the principle of **fairness** is to ensure that AI systems are non-discriminatory. Fairness consists of data fairness (requiring the use of non-biased datasets), design fairness (based on a meaningful system architecture), outcome fairness (consisting in non-discriminatory results) and implementation fairness (ensured by appropriately training end-users) [41] (13-22). **Accountability** should ensure answerability and auditability for AI systems. While answerability is concerned with the human responsibility for automation supported outcomes, auditability focuses on the mechanisms that should make designers and implementers of AI systems accountable and is strictly related to the justifiability of outcomes and their explicability [41] (23-25). **Sustainability** takes into account the transformative effects of the systems on individuals and the society and considers all possible kinds of impact on individual, society, future generations. Interestingly, sustainability also includes non-discrimination, although fairness (as non-discrimination) is reported as a separate principle. The sustainability of a project should be preventively assessed through a stakeholders' impact assessment, which shall assess the potential impact of the project on individual well-being and public welfare. Sustainability also includes safety and technical sustainability, meaning by "technically sustainable AI system" one that "is safe, accurate, reliable, secure, and robust" [41] (30). Finally, **transparency** should lead to explainability and justifiability of the systems and consists of two distinct dimensions: transparency of process and outcome transparency [41] (34-36).

These four principles do not stand, according to the ATI report, on an equal foot. Indeed, the report distinguishes between governing principles and qualities of AI systems: accountability and transparency belong to the first category and should inform the development and deployment process end-to-end. Fairness and sustainability, by contrast, are qualities of the AI systems [41] (13).

3. Identifying principles across the reports: towards a comprehensive and unified list of ethical principles

In this section, we systematize the principles mentioned in the four reports into coherent clusters. In doing so, we do not only take into account the labelling used, but also the core ethical content expressed by the principles as described by the reports. As will be seen, sometimes the reports use different terms for the same content, or denote with the same term principles with different ethical content.

Whenever a principle appears in at least two reports, we have taken it up in the unified principles list. We have also granted principle status if an ethical norm or value or aspect is taken up by at least one report as principle or as part of reflections on principle-overarching issues and is elaborated upon in at least one more report as part of the description of principles. In addition, if a report uses different, but synonymous or closely related terminology, we have conflated these issues under one heading or as a series of items. If a principle turns up only in the two AI or only in the two MB reports, we checked whether this principle could reasonably be applied in the other area as well.

For each suggested principle, we, firstly, broadly sketch its meaning as it emerges from our analysis of the reports and discuss comparatively how the principle is presented in the reports with regard to content and with regard to its status as principle, overarching value, or part of a description of a principle).

Secondly, as "findings", we shortly summarise the results of the comparative discussion, especially differences regarding the content of the principle in the reports, and take a stance on how to combine these understandings for a unified principles list.

If a principle is only mentioned in MB or AI reports, respectively, we include a section on "mutual learning" in which we argue whether the principle could be meaningfully applied in the other field of technology as well.

In a few cases, we have amended this methodology of interpretation and extraction with our own suggestions on how to develop further specific principles. These cases are added to the discussion and findings sections below as "suggestions".

In Section 3.10, we point to the ethical relevance of "conceptual framings" that are mentioned by some of the reports but cannot be dissolved into principles.

3.1. Human autonomy

3.1.1. Comparative discussion

A first principle that emerges in both fields, although with different modulations and in one case only implicitly, is respect for the human capacity to choose and alter values and convictions, to make decisions in light of those values and factual information, and thus to devise and follow one's own conception of a good life.

The AI HLEG guidelines adopt an individualistic conception of this **respect for human** autonomy or self-determination. They stress the importance of respecting human agency, conceived as the right to make informed decisions about AI systems and not to be subject to a decision made solely by an automated system. Moreover, they highlight the importance of ensuring human oversight on the functioning of AI systems [40] (15-16). The ATI report points in a similar direction stressing the importance for AI artefacts to support autonomy [41] (28). Both reports on AI are thus concerned with ensuring that AI technologies do not compromise the human capacity to make autonomous decisions about their (good) life.

The two reports on MB both do recognize respect for self-determination, albeit to a very different extent, and without using the term. The NCoB introduces the value of "equity", which is meant to cover and protect interests and value convictions of individuals as well as those of groups [35] (68). The NCoB thus understands respect for autonomy both individually and collectively, as relating to individual and group convictions and interests respectively. The NCoB argues that all those affected by a biotechnological application have an equal right to follow their interests and value commitments. This holds true, the NCoB asserts, also in those cases in which these commitments appear unfounded or irrational from another party's (or a purportedly "objective") perspective [35] (45).

The EGE, by contrast, mentions autonomy only (although affirmatively) when quoting documents such as the EU Charter of Fundamental Rights [34] (36). The EGE's own proposal does not include this principle. Nonetheless, one may regard autonomy as being included in "human dignity" [34] (39), and "intellectual property rights" [34] (45-47) may be viewed as a subcategory of autonomy rights, even though the EGE does not make either of those relations explicit.

3.1.2. Findings

Although not explicitly considered by all reports, and in one case captured under the differing label of "equity", respect for autonomy is introduced or referred to by all reports. It is an important ethical principle for AI and MB, and is represented as such in AI and MB reports.

Given the potentially widespread effects of MB applications on human lives, both with regard to place and time, the proposal of the NCoB to extend the autonomy principle to all stakeholders, as opposed to limiting this principle to scientists, holders of intellectual property rights or other specific group of stakeholders, and to apply the principle to both individuals and groups, appears reasonable. For the very same reason, it appears reasonable to adopt this wide understanding of respect for autonomy in AI as well. With applications in numerous areas, such as industry, health care, education, finance, the military, to name but a few, AI is set to influence the lives of all of us as individuals and members of social groups, now and in the future.

Respect for autonomy thus is relevant with regard to those individuals and groups who might be adversely affected by the outcomes of a technological application or who might have normative attitudes towards these outcomes that differ from a purportedly objective standard of best interests (or a similar concept), as well as for those individuals and groups who are involved in research, develop an application, or profit from it.

3.2. Individual and social well-being and prevention of harm

3.2.1. Comparative discussion

"Prevention of harm" and related principles, and "well-being", "public good" and "common good" can be understood as two sides of the same coin. They are both related to a standard of well-being as a factual state of affairs. While prevention of harm implies not to damage well-being, "well-being" (as an ethical principle) and "public good" require actively promoting and increasing a given level of welfare. Well-being and prevention of harm can be located at the level of individuals and spelled out in terms of physical, psychological and relational well-being and harm. They can also be located at the level of groups and societies and be explained in terms of economic growth, social equality, prosperity, social security, etc.

The ATI report introduces preventing harm as a principle. "Societal and environmental well-being" is subsumed under this principle (and, partly, fairness). Individual well-being is then mentioned as one component of "societal and environmental well-being" [41] (19).

The HLEG refers to ensuring "the physical and mental integrity of humans" as an element of the robustness of systems, which, in turn, is an element of the principle not to do harm [40] (16). In addition, the HLEG guidelines refer to serving "humanity and the common good" [40] (4) as the basic commitment that should govern all uses and applications of AI.

The EGE report contains no explicit reference to prevention of harm, well-being or public good. The closest it comes to this principle is by invoking "safety" and "biosafety", which can be understood, following the explanation of the EGE, as a specification of a principle to avoid inflicting harm on humans and the environment [34] (42, 48f).

The NCoB grants high level importance to the principle of (in the terminology of the NCoB) "public good". According to the NCoB, public good comprises a duty to increase human well-being as well as avoid doing harm. "Public good" is not among the NCoB principles sensu stricto, but describes the overarching norm towards which the public discourse ethics of the Council is supposed to be oriented as a whole.

3.2.2. Findings

Three reports include "prevention of harm" or related principles, or "well-being", "public good" or "common good". With the exception of the NCoB, the focus of these reports is on preventing harm. Nonetheless, although subordinately, a demand to maintain and increase well-being is taken up as well in the AI reports, especially as a demand to increase social well-being. Prevention of harm and well-being thus should be included in a unified list of principles. We have opted not to take a stance on the issue of prioritizing well-being and prevention of harm, but refer to both of them as one principle in the unified principles list.

HLEG and NCoB introduce "public good" and "common good", respectively, as single norms of the highest level of abstraction, under which all other principles can be subsumed. This may be a legitimate terminological choice. It entails, though, that public and common good cannot come into conflict with other principles. This appears to be counter-intuitive. As a current example, conflicts between, on the hand, individuals who reject being vaccinated, and, on the other hand, political and social actors aiming to safeguard the health of a population via vaccination in a pandemic appear to be best described as conflicts between respect for autonomy and the public good.

Many other cases of this kind can be imagined in the domains of AI and MB. For instance, with reference to the debate on human enhancement [42] (17), it is possible to imagine a future in which AI systems or MB applications could be used to enhance human well-being by, for instance, extending cognitive capacities or sensory skills. Again, if in such a scenario individuals or social groups reject or opt out of these technologies, the principle of autonomy would come into conflict with human well-being as a whole and the public or common good. For this reason, we have opted to include well-being and the common or public good as a principle on par with other principles.

3.2.3. Mutual learning

Both AI reports explicitly mention individual, physical and psychological well-being (or, "integrity"). Individual, physical and psychological well-being is not considered specifically in the MB reports. However, health and other applications of MB may very well threaten individual well-being. Given the tight and well-established rules and regulations of health applications, once an MB product reaches the stage of clinical trials, issues of individual well-being will be in the focus of attention, no matter whether or not such a principle is included in an emerging technology ethics list. Nonetheless, when development is still in the hands of biologists rather than medical researchers, early attention to these issues certainly make sense. Explicitly recognising individual well-being besides social well-being thus appears helpful for MB ethics as well.

3.3. Reliability, safety, and security

3.3.1. Comparative discussion

Another cluster of principles emerging in both fields is reliability, safety, and security. Basically, this cluster refers to the idea that engineered systems and artefacts should work as expected and designed for. It includes technical reliability of products and scientific soundness of the assumptions describing mechanisms, functions and results. It is thus related to justified expectations of users and other stakeholders that are to be met. It is also related to prevention of harm, especially individual physical and psychological harm, since one main expectation when using a system in the way it is designed for includes not being harmed in the course of doing so. Finally, the cluster also includes the demand to mitigate or eliminate the potential of a product or process for inflicting harm by being intentionally misused ("security").

The AI HLEG guidelines strongly support the call for technical reliability, highlighting the need for AI technologies to work as intended (including being robust and resilient against attacks), and adhere to good engineering practices (for instance by having a fall-back method in case the planned AI methods fail, by ensuring a low error rate and delivering reliable and reproducible results). This principle, as understood by the AI HLEG, also requires that the datasets selected for developing or training the AI systems are of good quality and adequate to the task [40] (16-17).

Issues of short-term reliability and scientific soundness of the underlying design and engineering assumptions do not get much attention in the MB reports. If safety is mentioned (the EGE has "biosafety"), the

focus is on long-term effects on ecosystems and, regarding especially the perspective of the NCoB, on prognostic uncertainties of the impact of applications on society and environment [34] (42), [35] (41-43).

Potential misuse, by contrast, is acknowledged in the field of AI as well as in MB, although not by all reports. Under the requirement of technical robustness and safety, the AI HLEG demands that AI systems must be resilient to attack and built in a way that prevents or mitigates opportunities for dual-use and malicious abuse [40] (27). However, both the ATI report and the AI HLEG guidelines do not to assign a prominent role among the ethics requirements to the prevention of misuse.

The EGE subsumes intentional misuse scenarios under the heading of "biosecurity", which is the well-established label for dangers of misuse in the ethical debate on biotechnologies (flanked by "biosafety" issues that are meant to cover unintended harms, as mentioned above) [34] (43f). The NCoB mentions misuse cases and scenarios [35] (43, 68f), but does not relate these cases to one specific principle.

3.3.2. Findings

Safety and reliability are introduced by both AI reports. Misuse is mentioned as well by both reports, although less prominently. Regarding the MB reports, misuse is mentioned, even prominently in the case of the EGE, whereas working safety and reliability of products and processes do not appear. Our findings thus suggest introducing safety and reliability as a principle, with misuse as subordinate part of it.

3.3.3. Mutual learning

One may assume that the omission of safety and reliability in the MB reports is due to the fact that the institutional environment in which development of AI takes place is very different from the institutional environment in the case of MB. Existing legislation related to MB, such as legal provisions regarding genetically modified organisms (GMO), may grant safe applications, as do research mainly conducted by qualified scientists and technicians in academic institutions and larger companies that provide internal oversight structures and quality management procedures. By contrast, the community of AI developers and their institutional background is much more heterogeneous and lacks standardized development protocols and verification means for AI systems. However, it should not go unnoticed that MB research does not presuppose huge investments and large institutional infrastructures. The biohacking scene bears witness to this fact. On this background, we conclude that systems reliability and safety is an important principle for MB as well.

3.3.4. Suggestions

Misuse (and "dual use of concern") is an important part of ethics and governance in MB. The low status the NCoB grants to this issue thus appears to be atypical for an MB ethics report. In addition, the technical equipment and know-how to run AI systems are widely accessible and the impact of classic AI misuse cases (in which an AI system designed to fulfil a beneficial purpose is taken as it is by an offender and, for example, used to manipulate democratic election) can be very severe. We thus propose that avoidance of misuse should play a greater role than granted by three of the four chosen reports in a unified list of ethics principles. We suggest incorporating security on par with systems reliability and safety in one principle.

3.4. Informational privacy

3.4.1. Comparative discussion

Another principle referred to in the AI reports, although in one case not as a separate principle, is **privacy** [40] (17), [41] (13, 28). In these reports, this principle is conceived mainly as informational privacy and data protection and covers issues related to data collection, data processing and access to data. Proper data collection procedures require that subjects consent to their use and are aware of the implications of their consent, while ethically suitable data processing activities ensure that data provided by subjects are used fairly, for instance by avoiding to extract information that can be used to discriminate against them. Finally, access to personal data shall be regulated and granted only to authorized and qualified personnel.

Issues of privacy do not appear in the ethics sections of the two MB reports. Related terms and issues, such as data protection, data safety, the right not to know, and informational self-determination, do not appear, either.

3.4.2. Findings

Privacy turns up in both AI reports and thus, according to our methodology, is to be included in a unified list of principles. Remarkably, the two MB reports do not mention this issue. This may be due to the fact that the ethical debate on molecular biotechnologies typically emphasizes applications in the sectors of food, agriculture, energy and other sectors not related to human health.

3.4.3. Mutual learning

We maintain, nonetheless, that issues of informational privacy can emerge in areas of application for MB technologies as well, most prominently in health applications. For example, research on MB compartments being used within the human body for diagnostic purposes is ongoing. Such systems might come up with incidental findings or they might uncover information about the very likely onset of severe, but as yet non-symptomatic diseases in later life. The ethical relevance of these scenarios has been long acknowledged in medical ethics, healthcare, and the governance and legislation concerning, especially, genetic testing [43]. Thus, informational privacy can be regarded as a useful addition to any list of principles intended to cover MB applications.

3.5. Transparency

3.5.1. Comparative discussion

A further principle, included in all reports, in one form or the other, is transparency. We take transparency, at first approximation, to require that stakeholder have access to relevant information. The terminology used across the reports is not consistent, though. For instance, both the HLEG Guidelines and the ATI report subsume under this principle aspects pertaining to accountability and vice versa [40] (18), [41] (34-36).

According to the AI reports, transparency requires that relevant information on the systems, broadly conceived, must be accessible to stakeholders, at least on request [40] (18), [41] (34-36). This includes, firstly, information regarding the development of the technological system. Methods, sources, and decisions regarding data acquisition and algorithm selection are supposed to be traceable and open for inspection. Secondly, this includes information on technological details of the final systems. Decisions taken by or with support of AI systems shall be explainable and not hidden in a black box system. Thirdly, transparency of interaction between humans and AI systems requires that humans must know that they are interacting with an AI system whenever this is the case. Transparency are, thus, once more an example of the focus of AI reports on the actual systems, their development and immediate impact.

In the case of MB, presumably because the reports emphasize long-term effects, transparency of system development and functioning is not mentioned. However, calls for transparency, accountability and democratic legitimization of governance structures and political decision-making concerning biotechnologies do figure prominently in the NCoB report. These procedures, according to the NCoB, need to be transparent and follow democratic standards, and accountabilities within this process need to be clear and revisable [35] (72). The EGE lists the development of procedures that ensure transparency (and access to information, democratic participation, and accountability and responsibility) as one of the "fundamental questions" MB raises [34] (40). This is not mirrored in the EGE's list of MB ethics principles and aspects, though.

3.5.2. Findings

Transparency is introduced in both AI reports and thus is to be included in the unified principles list. In addition, although the focus of transparency issues in the NCoB report is on governance rather than systems and products, transparency does turn up as well.

3.5.3. Mutual learning

We assume that the reason for not mentioning technical and interaction transparency in the BM reports derives from the fact that, with strict safety-related governance structures such as the GMO regulations in place, at least in Europe, the need to have access to information about technical details of a novel biotechnological product is not perceived to be as pressing. At the same time, the EU requirement to label products containing GMOs may be understood as a transparency requirement on par with the AI ethics demand to let users know when they are interacting with an AI system. It would make sense, then, to adopt the interaction transparency demand explicitly in the domain of MB as well, in line with and partly overlapping with the NCoB's concerns regarding different sets of values and inclusive and candid communication (cf. below).

Additionally, the demand for transparency regarding governance structures and decisional procedures in the MB reports mirrors the fact that the impact of biotechnologies stretches beyond a few individuals to whole societies and future generations. Accordingly, decision-making and governance concerning technology development and implementation should be accessible and revisable by society. The AI reports, by contrast, appear to conceptualize AI technologies rather narrowly, as technologies only affecting specific users or groups of users. However, given the potential impact of AI on the lives of all of us, and especially challenging cases such as automated behaviour of cars guided by AI, transparency of governance decisions concerning those systems would be a reasonable part of AI ethics principles as well.

3.5.4. Suggestions

Transparency can be understood, as in the ATI report, as a procedural, rather than substantial principle. One could argue, in line with this characterisation, that transparency is not to be seen as a principle in itself, but rather as a means to realise other values and principles. For instance, the requirement of making explicit whenever interaction with an AI system or a GMO is taking place can be understood as a way of ensuring human autonomy. Transparency of governance structures can be understood as being a prerequisite for accountability and democratic participation, and so on. In addition, as was shown, transparency covers very different ground. It encompasses transparency of the development process of an AI system, avoiding black box systems, displaying the artificiality of an AI system interacting with humans (and the GMO status of MB products, as argued above), as well as governance structures and decisions. One may conclude that transparency ought to be specified in more detail and be allocated to different, appropriate principles. We opted to list transparency as a standalone principle instead of scattering it among other principles, anyway. We take this to be a pragmatic choice, in line with the general emphasis of the AI reports on this issue, which ensures that transparency concerns are visible and cannot be overlooked during the ethics assessment exercise.

3.6. Accountability

3.6.1. Comparative discussion

Broadly speaking, accountability can be defined as the possibility to identify the individuals or institutions who have taken a given decision and to hold them accountable for it.

Both AI reports list accountability as a separate principle [40] (19-20), [41] (23-26). Key aspects attributed to this principle are auditability, accountability for trade-offs, and redress. Auditability refers to the possibility for oversight bodies or external experts to independently assess and review the development process and the functioning of a given technology. The second aspect of accountability requires that decision-makers are responsible for the decisions taken about trade-offs, while redress includes effective ad accessible mechanisms that can be activated in case of adversial consequences or impacts.

Regarding the MB reports, the EGE refers to "accountability and responsibility", together with transparency, as one of the "fundamental questions" raised by MB [34] (40). However, this point is not taken up in the EGE's list of MB ethics principles and aspects. The NCoB report includes accountability among its "procedural virtues". Accountability is explained as concerning governance structures and procedures. The public and social actors are supposed to be able to acknowledge and accept where responsibility for governance is located, and they should be able to revise the allocation of responsibility via democratic procedures [35] (69).

3.6.2. Findings

Accountability is mentioned as principle, or related to principles, in AI and MB reports. While AI reports focus on accountability concerning the design and development process of a system, the MB reports explain accountability in terms of governance structures, decisions, and decision-making procedures. Accountability with regard to both of these contexts thus should be taken up in a unified ethics principles list.

AI technologies have opened up the possibility that decision-making becomes a shared activity of humans and AI artifacts and is no longer an exclusive capability of human beings. However, while human beings are hold to be moral (and therefore accountable) agents, this is not the case for machines and artifacts. The need to fill in this gap makes accountability a particularly prominent issue in the ethics of AI.

3.6.3. Mutual learning

It is apparent that although the AI reports focus on short-term and single-technology related aspects of accountability, AI technologies do have a considerable impact on society as a whole, as well as long-term effects. We conclude, therefore, that long-term and governance aspects of accountability should be regarded as an important aspect in the field of the ethics of AI as well.

Conversely, although systems development accountability does not turn up in MB reports, applications of novel, genetically modified organisms in the environment may constitute a challenge similar to AI. To a certain degree, and over a longer period of time, synthetic organisms do become independent of their human designers, when they undergo evolutionary change and interact with other organisms in an ecosystem. This gives rise to the question where accountability of human designers may end. We conclude that systems-related accountability may be a relevant ethics principle for MB ethics as well.

3.7. Communication, participation, and democracy

3.7.1. Comparative discussion

Communication, participation and democracy are meant to ensure that, firstly, the public and stakeholders potentially affected by a new technology have access to all information needed in order to take a stance on the development, deployment, and use of a future or already existing technology in relation to their interests, preferences, and the public good. Secondly, it encompasses a bi-directional element aiming to ensure that the preferences and stances of affected stakeholders are taken seriously and can influence decisions about technology development and deployment, thus enabling stakeholders' active participation into decisions that potentially affect their lives. Thirdly, these abilities can be understood as part of a wider framework of a democratic distribution of power.

Transparency requirements regarding the functioning of a technology and transparency with respect to human-machine interaction partly overlap with communicative duties of scientists, developers and decision-makers towards users and the public. Communicative duties can be much wider, though. Furthermore, while the principles of

transparency and accountability mainly safeguard an on-demand and ex-post ability of stakeholders to reconstruct who took a decision, which decision, and when, communicative duties have a more proactive character. They refer not so much to the phase in which technologies have already been developed and are in use, but to the process prior to the decision to develop and deploy a given set of systems or technologies.

The AI HLEG mentions several aspects of communicative duties, although these are not listed as a separate category of principles. For instance, the AI HLEG guidelines mention (under the requirement of "diversity, non-discrimination and fairness") the need for involvement and participation of relevant stakeholders in the process of developing a novel system [40] (19). Moreover, the AI HLEG guidelines mention a demand to justify how and why certain technologies shall be deployed. As part of this justificatory process, relevant knowledge, especially regarding possible negative impacts of the new technologies and the trade-offs involved, shall be clearly and openly communicated. Finally, the AI HLEG guidelines mention protection of whistle-blowers as an additional requirement related to communicative duties. The AI HLEG report subsumes these last issues under the label of "accountability" [40] (20).

The AI HLEG mentions issues of democracy in the part of the guidelines concerned with the legality of AI systems and their compatibility with fundamental rights [40] (11). These considerations are not taken up in the ethical discussion.

With regard to ethics of molecular biotechnologies, the EGE report lists adequate participation of the public in fundamental issues of science and research among major challenges MB gives rise to [34] (39f). In addition, the EGE refers to "public involvement and science-society dialogue" as important parts of a governance framework for MB [34] (37). The EGE mentions that science and society dialogue ultimately involve issues of democracy and who is to decide what is to be allowed and what is not be allowed [34] (55).

The NCoB supplies a list of communicative virtues and principles that are meant to enable comprehensive non-biased public debates and decision-making on fostering and developing novel technologies [35] (68-71). These include "openness and inclusion", which parallels the EGEs reference to participation. The virtue principles also include "candor", a willingness to supply all relevant information, especially regarding the presentation of uncertainties and regarding alternative social or technological solutions, if they are available [35] (69f). Public reasoning, the NCoB claims, should be aiming at discovering common ground rather than confronting conflicting individual or group interests [35] (68). The NCoB adds that public engagement rests on basic democratic values [35] (82).

3.7.2. Findings

Three of our reports make reference to communication. The EGE lists this principle explicitly alongside participation. The NCoB splits communicative duties up into several virtue principles. This mirrors the general aim of the NCoB to supply a "discourse ethics" in which communication of course figures prominently, while it also results in the fact that the NCoB does not introduce a single principle of communication. The AI HLEG guidelines refer to aspects of science communication as specifications of the requirements of diversity and accountability. Both MB reports refer to participation as a principle (the NCoB has "openness and inclusion" as label). This overall picture warrants taking up communication and participation as a principle in the unified principles list.

3.7.3. Mutual learning

In the field of AI ethics, taking up participation as an explicit principle appears to be a sensible amendment. As mentioned above, given the societal impact of AI technologies such as AI assisted surveillance, a simple information model of communication will not be up to the task to do justice to the many differing impacts such a technology may can on the everyday life of social groups and individuals and on their respective preferences.

3.7.4. Suggestions

Both MB reports connect communication and participation to democratic procedures and values. The AI HLEG refers to democracy as part of legal considerations. They do also refer to the potential impact of AI systems on democratic electoral processes. However, they do not elaborate on these aspects and subsume them under the label of "societal and environmental well-being" [40] (19).

AI technologies can have adverse effects on the ability of individuals to make up their own mind, they can prevent them from expressing certain ideas and convictions, from participating in the public debate, and from participating in elections. Examples include filter bubbles and AI-supported disinformation campaigns, chilling effects induced by AI-supported surveillance (such as face recognition systems in public places or social media monitoring), and electoral manipulations.

Considering the impact that AI technologies can have on the very conditions of democratic processes, and given the reference of the two MB reports on democracy in the context of communication, we suggest that democracy should play a more prominent role among the unified principles. We thus suggest including it in a cluster together with communication and participation.

3.8. Justice, fairness, and non-discrimination

3.8.1. Comparative discussion

Issues related to justice and fairness can be regarded as oriented towards avoidance or mitigation of social discrimination, or, more generally, as the "fair" (including but not restricted to non-discriminatory) distribution of burdens and benefits of a technology.

Under the heading of "fairness", and in the case of the HLEG Guidelines in association with diversity and non-discrimination, the two AI reports stress in particular avoidance of bias and discrimination [40] (18-19) [41] (13-19). Potential bias is a prominent issue in the ethics of AI. Bias can arise as a consequence of historical discriminations that are reflected in the datasets used to develop and train the AI systems. The resulting AI systems can thus contribute to perpetuate historical inequalities, amplify existing prejudices and give them an appearance of "objectivity", for they can be interpreted as the result of computational practices instead of human judgement. Biases can also result from the technical design of the AI systems themselves, for instance from the way algorithms are programmed. Biases of the AI systems can thus cause discrimination by systematically producing results imbalanced against certain societal groups (examples include ranking women systematically less suitable for higher positions in jobs, and mistakenly assigning a higher risk of crime recidivism to black people [44,45,46]. Moreover, discrimination can also occur if some social groups are systematically prevented from using AI systems. This is why the HLEG on AI in its guidelines stresses the need to ensure that AI systems are designed in a way that makes them usable by everybody, irrespective of age, gender and abilities. Overall, these requirements aim to avoid reiterating and aggravating existing social discrimination, and to promote equal access to AI system via universal design.

Discussions of distributive justice are common in public debates on ethics of MB, especially issues of global justice [47]. The EGE includes justice among its central ethical aspects relevant for MB. On the account of the EGE, it comprises global justice as well as intergenerational justice and, accordingly, extends to environmental protection [34] (45). The NCoB calls for protection against unjust, harmful inequalities as well as to active commitment to those who are vulnerable or disadvantaged. The NcoB subsumes these two demands under the heading of "solidarity" and does not restrict this principle to issues of discrimination.

3.8.2. Findings

Both AI reports introduce issues of non-discrimination, in one case

using the term "fairness" only. The EGE refers to a broad, intergenerational and global principle of justice. The NCoB makes use of the term "solidarity" to refer to avoiding and actively ameliorating harmful, unjust inequalities between social groups. Terminological differences notwithstanding, issues of justice, fairness and non-discrimination thus clearly belong to a unified list of principles.

3.9. Sustainability

3.9.1. Comparative discussion

The principle of sustainability focuses on long-term effects of the use of natural resources. These resources are supposed to be used in such a way and to such a degree that their long-term availability is guaranteed. Sustainability is thus related to the well-being of future generations. In addition, sustainability may also be regarded to be based on the inherent value of non-human sentient or living beings. On this account, resources such as ecosystems should not only be protected as a resource for future generations, but also in order to acknowledge the inherent value of living nature.

In the field of AI, the HLEG guidelines highlight the need for a sustainable use of resources and energy consumption, as well as for impact assessment procedures that shall identify possible negative impacts and ways to mitigate them [40] (19). The ATI report, by contrast, recognises the principle of sustainability, but does not include environmental issues and focuses instead on the impact on individuals and communities [41] (26-30).

Regarding biotechnologies, the NCoB lists sustainability as one of its three "public values". The Council emphasises the importance of sustainable use of natural resources as part of responsibility for the well-being of future generations [35] (64). The EGE does not use the term "sustainability", but the biosafety issues the EGE mentions and the fact that the precautionary principle figures prominently (both with a focus on environmental impact) indicate that sustainability is an issue for the group [34] (42f).

3.9.2. Findings

Sustainability is a key principle of MB and AI alike. It is mentioned, implicitly or explicitly, in three reports. As was argued, until recently the principle of sustainability was underrepresented in AI ethics guidelines [26]. We note that, regarding the two AI reports chosen here, sustainability has entered the ethical debate in the field of AI as well.

3.10. Besides principles: conceptual framings

3.10.1. Comparative discussion

Both MB reports mention the ethical relevance of conceptual framings of technologies, their objects, and their prospects and applications. Taking account of an academic debate [48], the EGE points to the machine metaphor for living beings, pervasive in MB research, and its normative implications. Metaphors of this kind may induce profound changes in the understanding of the relation of humans to nature [33] (40f). The NCoB refers to "narratives" that may distort expectations. For example, if MB is depicted as "third industrial revolution", expectations may be grander than warranted by the facts [35] (35). These narratives, the NCoB continues, are spelled out in terms of metaphors such as the machine metaphor for living beings [35] (36).

Since the effects of these narratives and metaphors can be manifold and difficult to trace, and since they concern descriptions of states of affairs as well as values, ethical issues of this kind cannot easily be dissolved into ethics principles. None of the reports mentioning conceptual framings includes them as principle. We have therefore opted not to introduce these issues as a principle.

3.10.2. Mutual learning

Although neither the AI HLEG nor the ATI report discuss these aspects explicitly, the use of the anthropomorphic expression "intelligence" in reference to AI certainly plays a role in shaping the debate on AI ethics. Depicting AI systems as "intelligent" may misleadingly suggest that these systems possess a sort of judgment ("Urteilskraft" in the Kantian sense) or common sense comparable to humans [49]. This, in turn, may raise wrong expectations towards technological artefacts, and it might lead to premature acceptance of results and suggestions made by such systems, by the public and end users as well as developers and researchers. Attention to narratives and metaphors may thus be a valuable part of AI ethics as well.

4. Discussion and conclusion

Our analysis of the four ethics reports enabled us to compile a unified list of ethics principles. The principles are: autonomy; individual and social well-being and prevention of harm; reliability, safety and security; informational privacy; transparency; accountability; communication, participation and democracy; justice, fairness and non-discrimination; sustainability.

Our comparative approach enabled us to identify blind spots of the ethics principles as presented in the reports of each field of research respectively. Generally speaking, the field of ethics of AI can take advantage of the long-term focus and broad perspective on stakeholders adopted in the field of MB. MB ethics can profit from the AI focus on technical reliability of products and their immediate effects on users. Table 2 below summarizes the main amendments for a more comprehensive ethics framework in the area of AI and MB respectively.

Moreover, we posit that the unified list of ethics principles derived from the two technology fields can serve as a step forward towards identifying ethics principles relevant for emerging technologies in general. Table 3 below provides a comprehensive synthesis of the ethics principles we have identified and their elements and content.

Prior proposals for such a unified list of principles have often been based on ethics principles developed for specific fields of technology. For instance, the list put forth by Brey [6] is based on principles developed with a focus on IT-technologies. It introduces harms and risks, rights, justice, well-being and the common good as top-level principles. These top-level principles are then spelled out in terms of further subordinate principles. For example, "rights" comprises "autonomy", "human dignity", and other "basic human rights as specified in human rights declarations (e.g., to life, to have a fair trial, to vote, to receive an education, to pursue happiness, to seek asylum, to engage in peaceful protest, to practice one's religion, to work for anyone, to have a family, etc.)" An elaborated discussion of Brey's approach is beyond the scope of this paper and a matter of further debate. For the time being, we would like to point out that Brey's list does not mention ethical principles of transparency, accountability, and communication and participation. Prima facie, this point appears to count in favour of the comparative approach.

As another example, Gutmann [23] presents a list of five ethics principles which are supposed to be applicable to all emerging technologies (18). The principles stem from a report of the US Presidential Commission for the Study of Bioethical Issues on synthetic biology [22]. The list comprises five items: public beneficence, responsible stewardship, intellectual freedom and responsibility, democratic deliberation, and justice and fairness. As compared to Brey, this list is closer to our proposal. Nonetheless, it is apparent that the list does not include direct reference to transparency and accountability, which may be due to a lack of attention to the importance of accessible and revisable governance. In addition, autonomy turns up only as "intellectual freedom", which is to say as autonomy rights within research and technology development. Autonomy rights of other social actors are not mentioned. Again, this point indicates, we suppose, the merits of comparing ethics principles devised for different fields of technology in order to arrive at a comprehensive, unified list of principles for emerging technologies.

At the same time, the comparative approach suffers from a number of drawbacks as well. Firstly, if one follows a method of in-depth analysis

Table 2
How AI ethics principles can profit from MB principles and vice versa.

	Amending AI ethics principles	Amending MB ethics principles
Individual and social well-being, and prevention of harm		Well-being is understood as social well-being and "common good" in MB reports. Given the relevance of MB health applications, individual, physical and psychological well-being should be part of MB ethics principles as well.
Reliability, safety, and security		MB reports focus on long-term biosafety. Given the heterogenous disciplinary background of MB researchers and the activities of "biohackers", short-term reliability and safety of MB products should be recognized as well.
Informational privacy		Informational privacy is not part of MB ethics principles so far. Health applications may include MB diagnostic tools. In this context, informational privacy, the right not to know, and informational self-determination are important principles.
Transparency	Transparency is mainly understood as transparency concerning systems, their functioning, their interaction with humans, and the data used. AI applications can in many cases have an impact on whole societies (surveillance, automated driving). Hence, transparency of governance structures and decisions should be included.	Transparency is mainly understood as transparency of governance. Some MB products are subject to controversial public debate. Thus, MB product-related transparency (e.g. labelling) should be included.
Accountability	Accountability regarding harmful short-term effects of systems is recognized. Given possible long-term effects of AI systems that may affect whole societies, accountability regarding governance should be recognized as well.	Governance accountability is recognized. MB products, such as synthetic organisms, may have adverse long-term effects on the environment due to evolutionary change. Hence, product-related accountability should be included.
Communication, participation, and democracy	Science communication is understood mainly as informing the public and stakeholders. Participation is not mentioned. AI systems, such as AI assisted surveillance, may cause significant societal harm. Thus, bi-directional communication and participation should be part of AI ethics principles as well.	

Table 3
A unified list of ethics principles for AI, MB and beyond.

Principles	Elements
Human autonomy	– Respecting autonomous decisions of individuals regarding their own future. – Respecting autonomous decisions of social groups regarding their own future. – Respecting different individual and group convictions on what constitutes a good life. – Allowing individuals and groups to opt-out of dominant conceptions of the common good. – Ensuring human control of the decisions of technology systems. – Providing just rewards for individual and group efforts (via, e.g., intellectual property rights and free market conditions).
Individual and social well-being and prevention of harm	– Preventing physical and psychological harm to humans. – Preventing harm to the societal good and humanity. – Promoting physical and psychological well-being of humans. – Promoting the overall good of society and humanity.
Reliability, safety, and security	– Ensuring technical reliability of products. – Ensuring good engineering practices (such as fallback options). – Ensuring soundness of the scientific hypotheses underlying the product and its development (accuracy and reproducibility of results). – Ensuring robustness of products, including resilience to attacks. – Ensuring quality of data (such as non-biased and non-manipulated datasets) – Introducing measures to preclude misuse.
Informational privacy	– Respecting the right of individuals' to informational self-determination. – Deploying appropriate data governance (including, e.g., access rights to data) – Seeking informed consent before using personal data. – Protecting data.
Transparency	– Ensuring access to information for stakeholders and the public. – Ensuring traceability of development steps. – Ensuring transparency of decisions taken about how and when to deploy a given technology. – Ensuring explainability of products (including AI-based decisions). – Tagging the system when in interaction with humans (e.g. human-like AI systems; products in the focus of public debate, such as GMO) – Enabling openness of products for inspections and independent auditability. – Ensuring availability of information on governance structures and decision-making.
Accountability	– Implementing redress mechanisms in cases of mistakes or damages related to products. – Ensuring public acceptance of where responsibility for governance is located. – Providing mechanisms to enable the public to revise allocation of governance responsibility.
Communication, participation, and democracy	– Providing information on novel technologies and products and their development to the public and stakeholders. – Justifying how and why certain technologies shall be deployed. – Enabling the public and stakeholders to take a stance on technologies, products, and their development. – Directing the debate towards discovering common ground. – Ensuring the ability of the public and stakeholders to take part in technology

(*continued on next page*)

Table 3 (*continued*)

Principles	Elements
	development and governance according to democratic standards. – Ensuring that the basic conditions of democracy are not undermined.
Justice, fairness, and non-discrimination	– Ensuring global and trade justice. – Ensuring intergenerational justice. – Avoiding to introduce unjust, harmful inequalities. – Actively reducing harms to disadvantaged and vulnerable groups. – Ensuring fair, non-discriminatory collection and processing of data – Enabling equal access to products (via, e.g., universal design).
Sustainability	– Avoiding depletion of natural resources. – Avoiding negative impact on ecosystems. – Avoiding harm to living/sentient beings.

and interpretation, the number of reports that can be analysed is limited. This limits the validity of the results both with regard to the principles taken to represent a single technology domain and with regard to the unified principles. In addition, the number of emerging technologies that can be taken account of is limited. Comparison to other and more technologies may yield further, and better, results.

Similarly, we acknowledge that by having accepted the four selected reports as our starting point, we have inherited limitations of their approach, especially their selection of information. One of these limits is the lack of public and wide stakeholder involvement in devising the reports. Although the reports call for inclusion and communication regarding technology assessment, The EGE and the ATI reports did not deploy this demand with regard to their own activity. NCoB and AI HLEG have done a step into this direction by launching a public consultation on the first draft of their document. We do think that ethics guidelines could profit from adopting further procedural frameworks aiming at a stronger inclusion of different stakeholders [50].

We conclude that a comparative analysis of ethical principles related to different technology areas is a promising approach for developing a unified list of ethics principles for emerging technologies. We maintain that our findings could be furtherly refined by considering guidelines and reports of other technology domains and by increased stakeholder participation that more research and debate is called for in order to further advance this endeavour.

Funding

This work was supported by the 'Deutsche Forschungsgemeinschaft' (DFG) under Germany's Excellence Strategy [CIBSS – EXC2-189 – Project ID 390939984].

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the members of the Human Rights, Science and Technology Working Group of the Connecticut/Baden-Württemberg Human Rights Research Consortium and our colleagues at the Department of Medical Ethics and the History of Medicine, University of Freiburg, for discussing draft versions of this paper.

References

[1] R. Owen, P. Macnaghten, J. Stilgoe, Responsible research and innovation: from science in society to science for society, with society, Sci. Pub. Policy 39/6 (2012) 751–760.
[2] R. von Schomberg, Prospects for technology assessment in a framework of responsible research and innovation, in: M. Dusseldorp, R. Beecroft (Eds.), Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden, VS Verlag für Sozialwissenschaften, Wiesbaden, 2012, pp. 39–61.
[3] B.C. Stahl, J. Timmermans, C. Flick, Ethics of emerging information and communication technologies: on the implementation of responsible research and innovation, Sci. Pub. Policy 44/3 (2017) 369–381.
[4] A. Grunwald, Technology Assessment in Practice and Theory, Routledge, London/New York, 2018.
[5] T.A. Tran, T. Daim, A taxonomic review of methods and tools applied in technology assessment, Technol. Forecast. Soc. Change 75/9 (2008) 1396–1405.
[6] P.A.E. Brey, Anticipating Ethical Issues in Emerging IT, Ethics Inf. Technol. 14/4 (2012) 305–317.
[7] L. Floridi, A. Strait, Ethical Foresight Analysis: What It Is and Why It Is Needed? Minds Mach. 30 (2020) 77–97.
[8] European Commission, Horizon 2020 Work Programme 2014–2015. 16, Sci. Soc. (2015). Revised.
[9] Horizon 2020 Programme. Guidance, How to complete your ethics self-assessment (2019). Version 6.14.
[10] European Parliament and the Council, Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 (2013).
[11] L. Floridi, Translating principles into practices of digital ethics: five risks of being unethical, Philos. Technol. 32/2 (2019) 185–193.
[12] K.D. Clouser, B. Gert, A critique of principlism, J. Med. Philos. 15 (2) (1990) 219–236.
[13] E.D. Pellegrino, The metamorphosis of medical ethics. A 30-year retrospective, JAMA 269 (9) (1993) 1158–1162.
[14] J.H. Evans, A sociological account of the growth of principlism, Hastings Cent Rep. 30 (5) (2000) 31–38.
[15] OECD, Future technology trends. OECD Science, Technology and Innovation Outlook 2016, OECD Publishing, Paris, 2016, https://doi.org/10.1787/sti_in_outlook-2016-5-en retrieved from, https://www.oecd-ilibrary.org/docserver/sti_in_outlook-2016-5-en.pdf?expires=1641385317.
[16] R. van Est, D. Stemerding, V. Rerimassie, M. Rerimassie, J. Timmer, F. Brom, From Bio to NBIC - From Medical Practice to Daily Life, Rathenau Instituut, The Hague, 2014.
[17] B. Gordijn, Converging NBIC technologies for improving human performance: a critical assessment of the novelty and the prospects of the project, J. Law Med. Ethics 34/4 (2006) 726–732.
[18] G. Bennett, et al., From synthetic biology to biohacking: are we prepared? Nat. Biotechnol. 27 (12) (2009) 1109–1111.
[19] G.M. Church, et al., Realizing the potential of synthetic biology, Nat. Rev. Mol. Cell Biol. 15 (4) (2014) 289–294.
[20] T.L. Beauchamp, J.F. Childress, Principles of biomedical ethics, 7th ed edn., Oxford University Press, New York, 2013, p. 459.
[21] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, et al., AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations, Minds Mach. 28/4 (2018) 689–707.
[22] PCSBI, *New Directions: The Ethics of Synthetic Biology and Emerging Technologies* (Presidential Commission for the Study of Bioethical Issues) (2010).
[23] A. Gutmann, The ethics of synthetic biology: guiding principles for emerging technologies, Hastings Cent. Rep. 41/4 (2011) 17–22.
[24] A.M. Sismour, S.A. Benner, Synthetic biology, Expert Opin. Biol. Ther. 5 (11) (2005) 1409–1414.
[25] A. Jobin, M. Ienca, E. Vayena, The global landscape of AI ethics guidelines, Nat. Mach. Intell. 1/9 (2019) 389–399.
[26] Algorithm Watch (2020) AI Ethics Guidelines Global Inventory. Retrieved January 20, 2021, from https://inventory.algorithmwatch.org/about.
[27] UNESCO (2022), Recommendation on the Ethics of Artificial Intelligence, Adopted in November 2021. Retrieved June 2, 2022 from https://en.unesco.org/artificial-intelligence/ethics.
[28] N. Berberich, T. Nishida, S. Suzuki, Harmonizing artificial intelligence for social good, Philos. Technol. 33/4 (2020) 613–638.
[29] A. Gupta, V. Heath, AI ethics groups are repeating one of society classic mistakes, MIT Technol. Rev. (2020). Retrieved January 20, 2022, from, https://www.technologyreview.com/2020/09/14/1008323/ai-ethics-representation-artificial-intelligence-opinion/.
[30] P.R. Wolpe, K.S. Rommelfanger, Ethical principles for the use of human cellular biotechnologies, Nat. Biotechnol. 35 (2017) 1050.
[31] M. El Karoui, M. Hoyos-Flight, L. Fletcher, Future trends in synthetic biology—a report, Front. Bioeng. Biotechnol. 7 (175) (2019).
[32] A.S. Khalil, J.J. Collins, Synthetic biology: applications come of age, Nat. Rev. Genet. 11 (5) (2010) 367–379.
[33] W. Weber, M. Fussenegger, Engineering of synthetic mammalian gene networks, Chem. Biol. 16 (3) (2009) 287–297.
[34] EGE, Ethics of synthetic biology, in: The European Group of Ethics in Science and New Technologies to the European Commission, 25, European Commission, 2009 (ed.), *Opinion*.
[35] NCoB, Emerging biotechnologies. Technology, choice and the public good, London: Nuffield Council on Bioethics, 2012.

[36] William P Cheshire, Toward a common language of human dignity, Ethics Med. 18 (2) (2002) 7–10.
[37] M. Coeckelbergh, AI Ethics, The MIT Press, Cambridge, MA, 2020.
[38] H.-L.E.G. AI HLEG, A definition of Artificial Intelligence: main capabilities and scientific disciplines, High-Level Expert Group on Artificial Intelligence, 2019.
[39] T. Metzinger, Nehmt der Industrie die Ethik weg!, Der Tagesspiegel (2019), 08.04.2019. Retrieved January 20, 2022, from, https://www.tagesspiegel.de/politik/eu-ethikrichtlinien-fuer-kuenstliche-intelligenz-nehmt-der-industrie-die-ethik-weg/24195388.html.
[40] (2019) Ethics Guidelines for Trustworthy AI *(High-Level Expert Group on Artificial Intelligence)*.
[41] ATI, *Understanding artificial intelligence ethics and safety*. A guide for the responsible design and implementation of AI systems in the public sector, The Alan Turing Institute 2019; main author Leslie, D. (2019).
[42] E. Parens, Shaping Our Selves: On Technology, Flourishing, and a Habit of Thinking, Oxford University Press, New York, 2014, p. 216.
[43] P. Nicolás, Ethical and juridical issues of genetic testing: a review of the international regulation, Crit. Rev. Oncol. Hematol. 69/2 (2009) 98–107.
[44] S. Barocas, A.D. Selbst, Big Datas Disparate Impact, Calif. Law Rev. 104 (2016) 671–732.
[45] E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: can language models be too big?, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT 21, 2021, pp. 610–623.
[46] E. Orrù, Minimum Harm by Design. Reworking Privacy by Design to mitigate the risks of surveillance, in: R. Leenes, R. Van Brakel, S. Gutwirth, P. De Hert (Eds.), Computers, Privacy and Data Protection: Invisibilities and Infrastructures, Springer, Dordrecht, 2017, pp. 107–137.
[47] ETC, Extreme Genetic Engineering. An Introduction to Synthetic Biology, ETC Group, 2007.
[48] J. Boldt, O. Müller, Newtons of the leaves of grass, Nat. Biotechnol. 26 (4) (2008) 387–389.
[49] M. Mitchell, How do you teach a car that a snowman wont walk across the road? Aeon. (2019). Retrieved January 20, 2022, from, https://aeon.co/ideas/how-do-you-teach-a-car-that-a-snowman-wont-walk-across-the-road.
[50] K.J. Wakunuma, B.C. Stahl, Tomorrows ethics and todays response: an investigation into the ways information systems professionals perceive and address emerging ethical issues, Inf. Syst. Front. 16/3 (2014) 383–397.
[51] A. Buhmann, C. Fieseler, Towards a deliberative framework for responsible innovation in artificial intelligence, Technol. Soc. 64 (2021), 101475.
[52] S. Mohamed, M.-T. Png, W Isaac, Decolonial AI: decolonial theory as sociotechnical foresight in artificial intelligence, Philos. Technol. (2020).
[53] F. Russo, Digital technologies, ethical questions, and the need of an informational framework, Philos. Technol. 31/4 (2018) 655–667.
[54] G. Wellner, T. Rothman, Feminist AI: can we expect our AI systems to become feminist? Philos. Technol. 33/2 (2020) 191–205.
[55] B. Mittelstadt, Principles alone cannot guarantee ethical AI, Nat. Mach. Intell. 1/11 (2019) 501–507.
[56] S. Lo Piano, Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward, Hum. Soc. Sci. Commun. 7/1 (2020) 1–7. Palgrave.
[57] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, M. Srikumar, Principled Artificial Intelligence: mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI, Berkman Klein Center Research Publication No. 2020-1 (2020). Retrieved January 20, 2022, from, https://papers.ssrn.com/abstract=3518482.
[58] T. Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds Mach. 30/1 (2020) 99–120.