TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

Methodology for ethical analysis, scan results of existing ethical codes and guidelines

Deliverable 2.1

Draft version submitted to the European Commission for review

D2.1 Methodology for ethical analysis, scan results of existing ethical codes and guidelines			
Work Package	WP2		
Lead Partner	De Montfort University, DMU		
Author(s)	Sara Cannizzaro (DMU), Laurence Brooks (DMU), Kathleen Richardson (DMU)		
Contributor(s)	Steven Umbrello (TUD), Michael Bernstein (AIT), Laurynas Adomaitis (CEA)		
Due date	30/09/21		
Submitted date	29/09/2021		
Version number	1	Status	final

Project Information

Grant Agreement number	101006249
Start date	01/01/2021
Duration	36 months
Call identifier	H2020-SwafS-2020-1
Topic	SwafS-29-2020 - The ethics of technologies with high socio-economic impact
Instrument	CSA

Dissemination Level

PU: Public	☒
PP: Restricted to other programme participants (including the European Commission)	☐
RE: Restricted to a group specified by the consortium (including the European Commission)	☐
CO: Confidential, only for members of the consortium (including the European Commission)	☐

Quality Control

Reviewed by:	Review date:
Anais Resseguier (TRI)	21/09/2021
Lisa Tambornino (EUREC)	27/09/2021

Revision history

Version	Date	Description
0.1	15/09/2021	Initial draft
0.2	22/09/2021	Second draft
0.3	29/09/2021	Third draft

Keywords

Ethical analysis methodology, ethics codes, ethics guidelines, ethics frameworks, review

How to cite

If you are using this document in your own writing, our preferred citation is:

Cannizzaro, S., Brooks, L., Richardson, K., Umbrello, S., Bernstein, M., Adomaitis, L., (2021). *Methodology for ethical analysis, scan results of existing ethical codes and guidelines*. TechEthos Project Deliverable. Available at: www.techethos.eu.

The TechEthos Project

Short project summary

TechEthos is an EU-funded project that deals with the ethics of the new and emerging technologies anticipated to have high socio-economic impact. The project involves ten scientific partners and six science engagement organisations and runs from January 2021 to the end of 2023.

TechEthos aims to facilitate "ethics by design", namely, to bring ethical and societal values into the design and development of new and emerging technologies from the very beginning of the process. The project will produce operational ethics guidelines for three to four technologies for users such as researchers, research ethics committees and policy makers. To reconcile the needs of research and innovation and the concerns of society, the project will explore the awareness, acceptance and aspirations of academia, industry and the general public alike and reflect them in the guidelines.

Table of contents

List of tables

List of figures

Definitions and abbreviations

Table 1: List of Definitions

Term	Explanation
Climate Engineering	Climate engineering is a family of technologies that enables the modification of natural processes and human activities looking to address and mitigate climate change locally and globally.

Term	Explanation
Digital Extended Reality	Extended Reality refers to AI-powered digital technologies (hardware and software) capable of perceiving and processing human sensorial outputs, e.g., voice, gestures, language, movement, emotions and other elements of human communication, as well as responding to these types of signals by creating an extended visual, audio, linguistic or haptic digital environment for users.
Neurotechnologies	Neurotechnologies are technologies that aim at affecting and emulating human-brain capabilities and functions through artificial replacements or add-ons in a two-way interaction between the brain and the external environment or systems.
Ngram	Computational linguistics and probability; continuous sequence of n itemsfrom a given sample of text or speech

Table 2: List of Abbreviations

Term	Explanation
AI	Artificial Intelligence
ATE	Anticipatory Technology Ethics
BCI	Brain Computer Interfaces
BWC	Biological Weapons Convention
CWC	Chemical Weapons Convention
DBS	Deep Brain Stimulation
DoA	Description of Action
DTC	Direct to Consumer
eIA	Ethical Impact Assessment
ELSI	Ethical, Legal, and Social Implications
ESG	Earth System Governance
eTA	Ethical Technology Assessment
FAO	Food and Agriculture Organization
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
IEEE	Institute of Electrical and Electronics Engineers
mERA	mHealth Evidence Reporting and Assessment

Term	Explanation
MR	Mixed Reality
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
PC	Project Coordinator
PDMF	Precautionary Decision-Making Framework
PII	Personally Identifiable Information
RRI	Responsible Research and Innovation
RRI	Responsible Research and Innovation
SN	Social Networks
SRD	socially responsible design
SRM	Solar Radiation Management
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
VR	Virtual Reality
VSD	value-sensitive design
WHO	World Health Organization
WP	Work Package
WTO	World Trade Organisation
XR	Extended Reality

1.Executive Summary

- This report presents a review of three approaches to ethical analysis, ATE, eTA and Future Studies; these are ethical frameworks which were used across several projects on the ethics of new and emerging technologies (eg. ETICA, SHERPA, SIENNA, PANELFIT, REELER) and begins to identify the framework, which will address the three levels of ethical technology analysis (as discussed in ATE), which will be further developed in the following tasks of WP2.
- It also presents the results of a scan of existing ethical guidelines and frameworks on new and emerging technologies and their socio-economic impacts carried out as part of the TechEthos project.
- The scan of ethical documents identifies ethical issues associated with our selected socio-economic impact technologies. The ethical dimensions are expected to provide scope for development and critical reflection of technologies that we anticipate will be developed and deployed in Europe and worldwide in the next five to twenty years. The scan will be used to ensure that the ethics framework and guidelines developed by TechEthos will be relevant and applicable for a wide range of new and emerging technologies.
- The methodology for the ethical analysis will be broken down into different phases and focus on: (1) ethical codes, (2) ethical frameworks, (3) ethical guidelines. These will be identified and analysed for each of the technology families.

This report builds on WP1 D1.1 (Technology Families) and the consortium selected technology families. These are:

- Climate Engineering Technologies
- Digital Extended Reality
- Neurotechnologies

The scan of ethical guidelines is based on (i) desk analysis, taking advantage of existing updated ethical guidelines, policy, industry and non-governmental organisations and governmental at international, EU and national levels (ii) search for relevant codes related to the specific technology families using inclusionary/exclusionary criteria (iii) search documents with relevant keywords and (iii) an adapted mapping analysis approach.

The TechEthos scan of ethical guidelines is drawn from a novel approach in grouping and clustering families of technologies, based on the functions, applications, ethical and societal challenges addressed, and the identification of criteria for assessing potential socio-economic impacts of these technology families.

2.Introduction

2.1   Background

The world is changing, and ethical priorities are shaping how societies engage with and produce technologies. The horizon scanning activity (T1.1) has identified the three technology families that are the focus of the TechEthos project. This report contributes to an ethical overview which adds further to the conceptual and practical frameworks required to understand the technologies' high socio-economic impact. Moreover, as D2.1 (Review of current approaches to ethical Analysis and scan of existing codes and frameworks) will show, this is a complex task as there is significant variability in the specific technologies selected, as well as the infrastructure and contexts in which they have impacts. Therefore, the ethical consequences are speculative at times, as well as evidence based. There is also overlap between the technology families that will be reported on when relevant.

This report reviews three approaches to ethical analysis ATE, eTA and Future Studies, and explores the process and the result of a scan of ethical guidelines on new and emerging technologies and their socio-economic impacts carried out as part of the TechEthos project.

The review of approaches and the scan of existing ethical frames aims to identify moral and practical issues associated with the selected socio-economic impact technologies. This work will prepare the ground for further ethical analysis which will be developed in the following tasks of WP2.

These technologies are expected to have high value to societies of the future. Future forecasting is beset with multiple difficulties, and ultimately is a 'scientific' version of fortune telling. It is difficult to assess the various ways in which people, technology and the economy interact. One way of mapping is to use ethics as a perspective to assess, judge and examine the interrelationships between different phenomena. The people whose ethical work we scan as part of this deliverable are writing from expert positions, either as scientists and technologists who are directly producing the artefacts, or they are stakeholders with a vested interest in the success of these technologies.

3. Methodology for ethical analysis

3.1   Review of approaches to ethical analysis

This first section of the deliverable reports on a short review of approaches to ethical analysis that exist in the literature and as have previously been applied in a range of technology contexts. While this is not a comprehensive account, it aims to identify the key criteria in each approach.

3.1.1   Anticipatory Technology Ethics (ATE)

This approach focuses on emerging technologies from the perspective of trying to identify what is both good and bad about them. However, as these technologies are being developed, it is one thing to say what ethical issues are known, or can be reliably expected, but then there are also the ethical issues that will emerge over time as a consequence of use. Brey (2012a) reviews four approaches to technology assessment focused on ethics, namely ethical Technology Assessment (eTA) (Palm and Hansson, 2006), ethical Impact Assessment (eIA) (Wright, 2011), techno-ethical scenarios (Boenink et al., 2010), ETICA approach (Stahl, 2011). Based on his analysis of these, Brey proposes a fifth approach, ATE, which he says has "the potential to meet all the criteria that a sound approach to ethical analysis of emerging technologies should have" (Brey, 2012a, p309).

ATE has three levels of ethical analysis: technology, artifact and application level. It then defines what are called 'objects of ethical analysis' for each of these levels, as properties or processes that might lead to ethical issues.

Table 3: ATE levels

Technology analysis:	Consider the impact of the technology independent of any artifacts or applications
Artifact analysis:	Consider the physical configuration of the technology, which, when operated in a proper manner produces the desired result.
Application Analysis:	Analyse the application of the technology within a specific context.

One of the issues for the early stages of ATE is how to identify the appropriate ethical values to be mapped with the specific technology. Brey (2012b) proposes an ethics checklist (see Table 4), which encompasses a range of ethical values and principles, based on ones that have been seen in earlier ethical approaches and commonly found within society.

Table 4: The anticipatory technology ethics checklist (Brey, 2012b)

Harms and risks	Rights
- Health and bodily harm - Pain and suffering - Psychological harm - Harm to human capabilities - Environmental harm - Harms to society	- Freedom - Freedom of movement - Freedom of speech and expression - Freedom of assembly - Autonomy - Ability to think one's own thoughts and form one's own opinions - Ability to make one's own choices

	- Responsibility and accountability
	- Informed consent
	o Human dignity
	o Privacy
	    - - Information privacy
	    - - Bodily privacy
	    - - Relational privacy
	o Property
	    - - Right to property
	    - - Intellectual property rights
	o 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.)
	o Animal rights and animal welfare

Justice (distributive) o Just distribution of primary goods, capabilities, risks and hazards o Nondiscrimination and equal treatment relative to age, gender, sexual orientation, social class, race, ethnicity, religion, disability, etc. o North–south justice o Intergenerational justice o Social inclusion	**Well-being and the common good** o Supportive of happiness, health, knowledge, wisdom, virtue, friendship, trust, achievement, desire-fulfillment, and transcendent meaning o Supportive of vital social institutions and structures o Supportive of democracy and democratic institutions o Supportive of culture and cultural diversity

Munoko et. al. (2019) summarise the 5 steps for the researcher to follow in ATE as:

- "First, at the *technology level*, the researcher considers the features of the technology of ethical concern, independent of its current or potential use. This level involves the identification of the inherent and consequential risks of the technology.
- Secondly, at the *artifact level*, the researcher considers the "physical configuration that, when operated in the proper manner and the proper environment, produces the desired result." At this level, the researcher focuses on the artifacts independent of their actual applications and identifies the risks associated with the intended use of the artifacts.
- Third, at the *application level*, the actual use of an emerging technology's artifact is studied. At this level, the researcher considers the unintended consequences for the users of the applications and other stakeholders.
- Fourth, the researcher evaluates the potential importance of the issues identified.
- Finally, the fifth part of the ATE framework is optional, where the researcher can design a feedback stage.
- There are additional optional stages beyond the fifth step. One optional stage is the responsibility assignment stage, where "moral responsibilities are assigned to relevant actors for ethical outcomes at the artifact and application levels." Another optional stage is the governance stage, which provides policy recommendations."

Munoko et. Al. (2019) then combine ATE with the ETICA approach (Stahl, 2011), as they feel that each of the methods, while closely linked, contributes something that the other does not.

More recently, ATE has been cited as one example of 'technology oriented assessment methods', including eTA, eIA, as well as value-sensitive design (VSD), privacy for design, socially responsible design (SRD), eco-design, ethics by design (Gurzawska, 2021).

One critique of ATE is that trying to predict what might be the impact and outcomes of emerging technologies, will be problematic, as until people actually start to use those technologies it is difficult to recognise what might be the unintended and emergent properties which will be seen. However, that is not to say that likely outcomes cannot be conceptualised and recognised, within a framework such as ATE.

3.1.2   Ethical Technology Assessment (eTA)

Ethical Technology Assessment (eTA) arises out of recognition of the long-term consequences of technology on society. As a field it has both a conceptual, as well as a practical component about how to incorporate ethics into the process of technological practice, not as an 'add-on' but inherent in the process. As a methodology, it is developed in conjunction with developers and has a 'continuous dialogue rather than a single evaluation at a specific point in time' (Palm and Hansson 2006: 543).

Palm and Hansson (2006) identify nine ethical aspects as critical aspects of technological impact including: 1. Dissemination and use of information, 2. Control, influence and power, 3. Impact on social contact patterns, 4. Privacy, 5. Sustainability, 6. Human reproduction, 7. Gender, minorities and justice and 8. Impact on human values. These are summarised as:

1. **Dissemination and use of information:** Use of technologies give rise to new patterns for the dissemination of information
2. **Control, influence and power:** There are many historical examples of how technological change has led to changes in the distribution of control and influence, not least on workplaces.
3. **Impact on social contact patterns:** Communication technologies such as the telegraph, telephone, radio, TV, Internet and cellular phone have affected the way people establish contacts, meet, and communicate.
4. **Privacy:** As a consequence of new and more sophisticated means for identifying and collecting different types of information about individuals, private spaces where individuals may remain free from intrusion, seem to diminish.
5. **Sustainability:** It has been increasingly recognized that the decisions we make today should be defensible also in relation to coming generations. New technologies may affect all three sustainability dimensions through their influence on economical, social, and ecological development.
6. **Human reproduction:** Some of the most blatant clashes between on the one hand social norms and moral values and on the other hand technological innovations, have taken place within the field of reproductive technology.
7. **Gender influence and power:** The advantages and disadvantages of technologies are often unevenly distributed between women and men. New technology often changes the relationship between nations and in particular between the developed and the developing world.
8. **Impact on human values:** There are many ways in which technological development affects the way we live, the way we understand ourselves and our moral values and principles.

eTA grew out of Technology Assessments (TA) – a framework used for the first time in 1966 in the US leading to a set of practices that aimed to identify technological consequences. eTA is an outgrowth of this earlier project - but its advocates argued that ethics must be integrated into the process of development, manufacture, and use of technological artefacts.

TA is not without its critics and was challenged as inefficient due to the way it narrowly shaped the assessments – for example, a focus on European nations over developing ones (ibid 546). In an

attempt to address the problems in TA, a new methodology called participatory Technology Assessment (pTA) emerged in the 1980s – this methodology could be put to use in contexts other than Europe and North America, and could be responsive to local issues, resource implications, and challenges by incorporating a flexible model into its practices. A range of other techniques emerged out of TA, in addition to pTA there was *Constructive Technology Assessment (CTA), Innovative Technology Assessment (ITA) and Health Technology Assessment (HTA)* – modes of assessment that developed in response to specific problems but without a prioritising of ethics. The medical fields were among the first to incorporate ethics, as the practices and new techniques were reshaping issues of human autonomy, privacy, the family, and reproduction.

Ethical reflections went beyond examining the impacts on humans, but also began to question the validity and need of the technology itself by "recasting the way problems are defined, by exploring the interrelationship of the technical and non-technical issues, and by analyzing technology itself as problematic" (Housemakers and Henk cited in Palm and Hansson 2006: 548).

Since these early experiments in integrating ethical assessments into technological processes, an effort that was either minimized as unimportant, or critically rejected by technologists as a barrier to innovation and progress, the argument for ethics has now been seen by governments, particularly the European Union, to be a central part of the practice.

As an approach, selection criteria are established about what areas of technological impact should be prioritised (see nine areas identified above). This has led to others to criticise its 'checklist approach' and predefining the ethical issues they write "This checklist reinforces a TA method in which the potential ethical implications of new technologies are evaluated according to given, fixed ethical principles and rules." (Kiran, Oudshoorn and Verbeek, 2015: 5).

Drawing on the field of Science and Technology Studies (STS), Kiran et. al. (2015) suggest ethics is co-terminus with societal development as 'co-evolutionary' changing underlying normative (what is considered the norm) judgements. Instead, they advocate 'technological mediation' where "Rather than locating human beings and technological artifacts in two separate domains – the domains of subjects and objects – this approach considers technology to be a medium for human experiences and practices." (ibid 4).

In this account, technologies act as interpretive and mediating devices for humans who need to engage with them to make sense of the world – an example used is the thermometer, whose reading indicates a scientific temperature that is not a phenomenological experience of heat or cold experienced by a human body– and as such, this work evolves out of the post-phenomenological work of Don Ihde (as referenced in Kiran, Oudshoorn and Verbeek, 2015).

Recognising the role of technological artefacts in shaping human knowledge about themselves and their world, inadvertently humans are subjected to 'behaviors and norms scripted by technology' (Kiran, Oudshoorn and Verbeek, 2015: 7).

eTA approaches depend significantly on what underlying model of human and technology is used – as Boer, Hoek and Kudina (2017) explain, the post-phenomenological approach to ethics through technologies shapes the future in particularl ways, they write "Read in this strong way, the proto-ethics of mapping phenomenological normativities is already an explication of a specific relation with the future within which potential actions are already assumed." (ibid). They instead propose a 'proto-ethics' that must include how the ethical practices are shaping the ethical outcomes. They identify weak and strong approaches to technological mediation. Weak approaches use ethics to complement TA approaches, while strong approaches demonstrate how 'phenomenological and existential

normativities go hand in hand' and thus in the latter approach, the technological user 'continuously discovers her/his own needs and expectations and is not assessing something external'.

In summary, eTA is a contested domain, as the method itself is open to criticism of how it cannot escape from the normative practices and it aims to disclose its own assumptions in the process.

3.1.3   Future Ethics

Prediction, foreseeing of the future is a key feature of all human cultures and was traditionally expressed by oracles, and augurs who could gift the future in the present. In modern scientific societies, prediction moved from the professions of clairvoyants, fortune tellers and prophets to professionals, academics who would develop techniques and methodologies for 'seeing the future'. A dictionary definition of the future is 'going or expected to happen or be or become' (Oxford English Dictionary cited in Sardar 2010: 178). The term 'futurology' was first used by Flechteim in 1966 with the publication of History and Futurology (Flechteim cited in Sardar 2009: 178) a new field that would explore the 'destiny' of humans. He regarded the subject as a branch of 'historical sociology'. In socio-technical capitalism societies, technologies are reshaping the present, offering up new possibilities, and problems. For this reason, scholars have begun to regard the current period of geological time as the *Anthropocene*, an epoch shaped by human activity, it "describe(s) a connection that reaches back into the past and far into the future' (Schwägerl cited in Marak 2019:19).

Future Studies emerges as an interdisciplinary field, recognising that the 'future' is not produced by one agent, but a number of intersecting, often colliding and reacting processes. A critical problem for it is the role of *time* – not understood as linear and singular but, Schneider (2019) explains it as the future seen as an outcome of gestures and properly studied as 'interval crossers' and 'interval openers' (147).

Future Studies also accounts for the role of imagination, and 'the imaginary as resources for (re-) shaping our world and imagining new relations' and prioritising the role that stories play in constructing human existence (Spengler 2019: 168).

Future Studies goes beyond prediction, as it aims to shape the future according to principles and values that are important to humans. But what is the future – is it anytime that is beyond the present, or a place that is always shaped by fictional imaginaries and any prediction must consequently be partly, a work of fiction. Moreover, artists, including novelists have shaped future predictions, from Issaac Asimov to Arthur C. Clarke (Potts 2018). Science fiction writer Ursula Le Quin warned against calling upon artists to predict the future, as she claimed they do the opposite – they tell lies (2016 [1969]. Le Quin also notes the struggles she had to write believable female characters into her science fiction in a male dominated field - as the human is often reproduced as the *default male* - woman has to be explicitly stated if they are to be included in future forecasts as noted by feminist writer Caroline Criado Perez in her book *Invisible Women: Exposing Data Bias in a World Designed by Me*n (2019).

Sociologist Zygmunt Bauman noted that the future is not always a desired goal, and he coined the term "*retrotopia*" as an umbrella term for those movements and trends that seek to get back to something, rather than moving somewhere else (cited in Paul 2019). Hence ideas of the future are intrinsically connected to the past and present, imagined and factual, as opportunities, and destruction are feasible outcomes of any process.

Future Studies is not without its critics, for to have a future must imply a desired or imagined state of existence – calling into question who decides this future? Who is left out or excluded from future imaginings? The question is whether technology innovation is the solution to the problems developed

in tech-capitalist societies? Technology, is the engine of capitalism innovation, opening up the possibilities of creating new products, processes and practices, underlying a belief in unfettered creativity and flexibility of the human species to adapt to any technologically inspired living arrangement.

Hojer and Mattsson (2000) work in transport research is useful for TechEthos, as they identified four critical problems with Future Studies approaches: 1) identifying 'cyclic behaviour in socio-technical changes'; 2) viewing one technology to be crucially reliant on the development of another (in their case it was transport and communication that entangled and connected), 3) interrogating basic assumptions about a field (in their case it was the 'hypothesis of constant travel time' as a stable), and 4) human and resource relationships (613). The future is a 'fiction' of sorts, shaped by practices, ideas and, extrapolated into some undefined future point – problematically producing a determinism – if this, then that – view. Moreover, they suggest that 'backcasting' as an alternative and better predictor than 'forecasting' in cases where future scenarios are seen as detrimental, and harmful. Sadar prefers the term 'alternative futures' due to the possibility of plurality, identity crises and meaning (2009).

Ethically speaking, the 'future', if it exists at all, is a contested domain, heterogenous, and diverse, while ethics proposes a set of standards to be recognised and incorporated into technological practices and artefacts. Artificial Intelligence (AI) is a case in point, with a past littered with inaccurate accounts – and yet evidence of failed predictions are passed over, as new ones form and develop (Sundvall 2019).

3.1.4   Summary

What all these approaches to ethical analysis show us is that it is difficult to predict the future. However, as techniques and approaches they each demonstrate that it is possible to develop some guidance on how to assess the possible ethical issues associated with a specific technology, so that developers and users may reflect on this and potentially incorporate those reflections into their design, development and use. Given the importance of these issues, TechEthos proposes to further develop its approach to ethical analysis in Task 2.2, using the three levels of ethical analysis (from ATE) as guidance.

3.2   Literature search strategy for the scan of ethical documents

In the second part of this report, we provided the results of a scan of ethical documents. The methodology for this literature scan was constructed using a mixed method approach. As a first step, we identified published reports, academic journal articles, books, and working papers that examined guidelines, ethical codes, codes of conduct, and governance frameworks as used within climate engineering, digital extended reality and neurotechnologies.

The key terms we used are:

- 'ethical codes'
- 'ethical frameworks'
- 'ethical guidelines'

Once we retrieved the above ethical frames (codes, frameworks and guidelines), as a second step we scanned the results further based on a selection of a number of fundamental ethical principles extracted from the consortium's guiding categories (Table 3), based in turn on Brey (2020b).The scan

results of ethical principles for each technology family varied considerably and different results were returned.

Table 5: A selection of Ethical Principles and Concerns based on Brey (2020b)

Fundamental principles
Impact on: • Human rights • Freedom • Autonomy • Integrity • Responsibility • Privacy • Security

The databases we searched included JStore, Google scholar, ACM Digital Library, IEEE Xplore Digital Library, AIS eLibrary, and we also carried out a general search on Google where we often found reports from companies, or organisations that are traditionally excluded from academic databases.

Our aim was to obtain a set of documents, comprising of both published academic literature and grey literature from either industry, government, non-academic and non-governmental (NGO) research and policy organisations that would have ethical guidelines, codes and frameworks as a key content in their text. By grey literature we intend "That which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers" (Greylit 2021). We did not seek to include texts which mentioned ethical principles in general without reference to specific guidelines, codes and texts.

To this end, we considered carrying out a comprehensive literature review search to assess whether this method would provide us *both* with a sufficient quantity *and* variety of relevant sources. We aimed to gather at least 20 sources per each of the technology family (Climate Engineering and Neurotechnologies), and 30 sources for extended Digital Reality given that we expected a higher quantity of documents to be associated to this tech family.

The search algorithm we used was:

> ("Abstract":ethic*) AND ("Abstract":guideline) AND ("All Metadata":Natural Language Processing OR "All Metadata":NLP)

That is:

1. 'ethic*' - which encapsulates ethics and all terms with ethics included – *within* ABSTRACT (this key term had to be present in the abstract of the document)
2. 'Guideline/Framework/Code' - *within* ABSTRACT and/OR author KEYWORDS
3. Technology family or specific technology type (in the case of extended digital reality) e.g. 'natural language processing' OR NLP – *within* ABSTRACT (this key term also had to be present in the abstract of the document)

The keyword 'ethic*' would capture documents containing all key terms related to ethics, such as 'ethics', 'ethical', 'ethic' at once, without needing to perform separate searches for each of the terms.
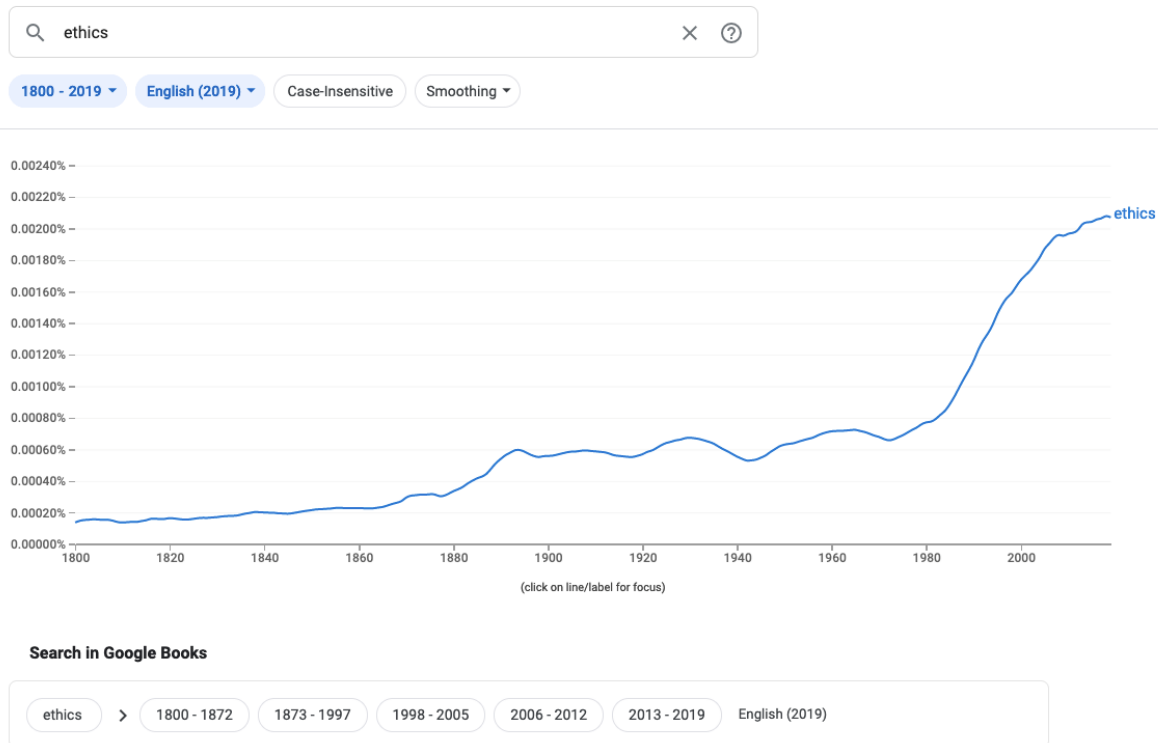
Figure 1: Ethics Google Ngram View Chart

Through this search, we wanted the specific technology type and the keywords 'guideline/code/framework' to be present in the abstract or at least the keywords, to ensure we would obtain documents that were specifically *about* ethical guidelines for the technology families and their specific technologies. However, we found that the systematic search of strictly academic databases was not producing enough relevant results, and was omitting grey literature. Hence it is where ethical reflections on emerging and established technology are also present and produced from a variety of stakeholders. An example of grey literature would be produced by businesses wishing to market their products as 'responsible' to attract a specific target of ethically minded consumers. The business in question would be looking to show adherence to or even produce their own ethical guidelines, codes and frameworks. For example, see Accenture's recent report on *Responsible AI: From principles to practice* (2021). Hence in order to capture these instances of ethical efforts within emerging technologies, we complemented the search with other search avenues such as Google Scholar and Google. These present the benefit over academic databased of drawing on a wider pool of sources but of course also present the issue their algorithm developed for commercial purposes not being transparent. There is a growing trend, however in science, of carrying out systematic literature searches including the results from Google Scholar, as for example shown in Seid et al (2018). We therefore settled for our literature search to include variety, relevance, comprehensiveness as its key methodological feature, but not representativeness of academically verifiable scientific publications.

3.3   Create a Zotero library

The results of the search were saved and imported into Zotero, an open-source reference management software to manage bibliographic data and research materials (Zotero 2021). We saved more than the 20 sources for each of the technology families for scrutiny of relevance, that is, to assess whether ethical guidelines, codes and framework were being foregrounded in the documents,

Extended digital reality was a too general search term to provide relevant results, hence it was split to include specific technology types as specified on the technology family factsheets (output from WP1), The types include AI, Augmented Reality, Digital Twins, Distributed Clouds, Edge Computing, NLP.
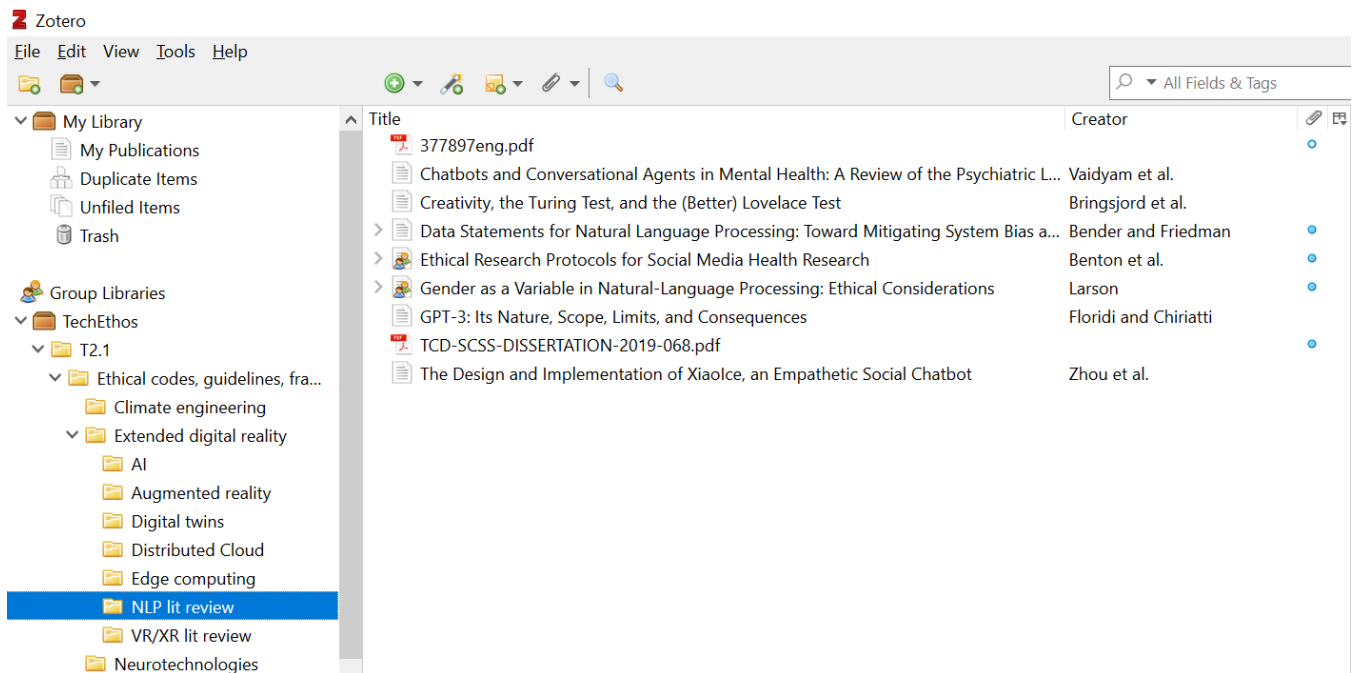

Figure 2: Zotero Library collection

Zotero was also chosen given the avenue it provides for sharing libraries of sources with other users, and hence as a tool to foster collaboration across the consortium.

3.4   Definitions of ethics guidelines/frameworks/codes

We sought to identify relevant ethics guidelines/frameworks/codes within the selected sources (as saved on our shared Zotero library) which we refer to as 'literature scanning' in this report. We note that the terms guidelines/frameworks/codes were used interchangeably in the literature. We also understand that guidelines/frameworks/codes can indeed be interrelated to each other in a complex manner, sometime hierarchically (for example codes and guidelines are considered by some as components of frameworks), hence are not strictly reducible to paradigmatic, self-contained definitions. However, for the purpose of this scanning exercise we did not aim to delineate such interrelations nor the hierarchical levels to which guidelines/frameworks/codes pertain since this would constitute a deeper level of analysis. For the scan, we utilised a technique aimed at *detecting* these ethical frames as they occur in the literature, not as they interrelate. With this in mind, we therefore identify the main difference between these terms to lie in their level of generality i.e. ethical codes have a narrower and more specific focus and guidelines have a broader scope, with frameworks laying somewhere in the middle in terms of level of generality. Below we capture and articulate further the distinction amongst these terms based on the example set by Rothenberg et al. (2019). Hence, we generated definitions of these terms with the purpose of defining in a clear-cut manner what constitute ethical codes, guidelines and frameworks.

Ethical codes

Ethical codes set forth responsibilities to which individuals and groups or organisations hold themselves to account. Compliance with codes may be enforced with socially mediated consequences

for non-compliance or rewards for compliance. Related to emerging technologies, ethical codes elevate individual responsibility to promote desirable and/or minimize undesirable developments in the field.

- Target: individuals, groups, organisations
- Proximal Goal[1]: enhanced responsible behaviour in the field
- Distal Goal: enhanced desirable and reduced undesirable outcomes of activities in the field
- Compliance: determined by social pressures; in rare cases also formal sanction
- Scope: may be unique to individuals/organizations or shared across many entities

Ethical frameworks

Ethical frameworks set forth general or specific principles to which countries, organizations, or research communities hold themselves to account. Frameworks arise in otherwise unregulated situations where groups of actors seek to alter the development trajectory of a field. Compliance with frameworks may be enforced with socially mediated consequences for non-compliance or rewards for compliance. Related to emerging technologies, ethical frameworks seek to coordinate alignments of the behaviour of collectives of individuals to promote desirable and/or minimize undesirable developments in the field.

- Target: countries; organizations; research and innovation communities
- Proximal Goal: enhanced coordination of responsible behaviour by disparate groups of actors in the field
- Distal Goal: enhanced desirable and reduced undesirable outcomes of activities in the field
- Compliance: determined by social pressures; in rare cases also formal sanction
- Scope: shared across many entities

Ethical Guidelines

Ethical guidelines collect general or specific principles specifying how a technology or field ought to develop. Guidelines may be generated through concerted collective action of individuals or organizations. Compliance is not usually considered with guidelines. Related to emerging technologies, ethical guidelines propose development pathways intended to enhance desirable and/or minimize undesirable outcomes of a field.

- Target: research and innovation pathway of a technology or research area
- Proximal Goal: agreement on responsible directions for a technology or research area
- Distal Goal: enhanced desirable and reduced undesirable outcomes of a field
- Compliance: not usually considered
- Scope: shared across many entities

3.5   Mapping

The strategy that this literature follows was set by Rothenberg et al (2019: 4); their review of ethical guidelines of AI extracted common guidelines from a select sample of relevant literature, grouping

---

[1] Proximal refers to objectives to reach in the short term which have higher probability but lower value. Distal refers to objectives to be achieved in the longer term which have lower probability but higher value. People are more likely to persist and achieve distal goals if they are connected with proximal goals.

them based on the ethical principle that underwrite them. They included the type of organization issuing the guideline and a definition for each.

Table 5: 6Table showing example of sample based on selected sources within the neurotechnologies technology family

Neurotechnologies: ethical guidelines, codes, frameworks and issues (based on Rothenberg et al 2019)			
Guideline	Type of organisation	Definition	Extract of source guideline
Ethical code	Academia	Ethical codes set forth responsibilities to which individuals and groups or organisations hold themselves to account.	...professional self-regulation […] should start within a company, institution or other work unit with a code of ethics or set of clearly articulated principles to which leadership adheres... (Chang et al 2019)
Ethical frameworks	Academia, other research organisation	Ethical frameworks set forth general or specific principles to which countries, organizations, or research communities hold themselves to account.	Australia currently lacks a clear regulatory framework for ensuring that individuals are informed about how their data are captured, stored, analyzed, and shared (Australian Brain Alliance 2019)
			The degree of perturbation of advanced neurotechnology on the current ethical legal framework is quantitatively higher than non-computational techniques (Ienca and Andorno 2017). Key ethical concerns that arose from the Brain/MINDS project include ethical standards concerning clinical data collection (...). The neuroethical framework constructed by the preceding national brain projects ... in Japan needed to be extended (Sadato et al 2019).
Ethical guidelines	Academia, other research organisation	Ethical guidelines collect general or specific principles specifying how a technology or field ought to develop	The clinical research organizing team has also created guidance for any necessary modifications needed in ethical protocols due to revisions of research guidelines (Sadato et al 2019).

Neurotechnologies: ethical guidelines, codes, frameworks and issues (based on Rothenberg et al 2019)			
Guideline	Type of organisation	Definition	Extract of source guideline
			To develop national guidelines for responsible neuroinnovation to assist neuroscientists, engineers, and developers to translate research into effective and ethical products. (ABA 2019).

4. Scan results of existing ethical codes, guidelines, frameworks and principles

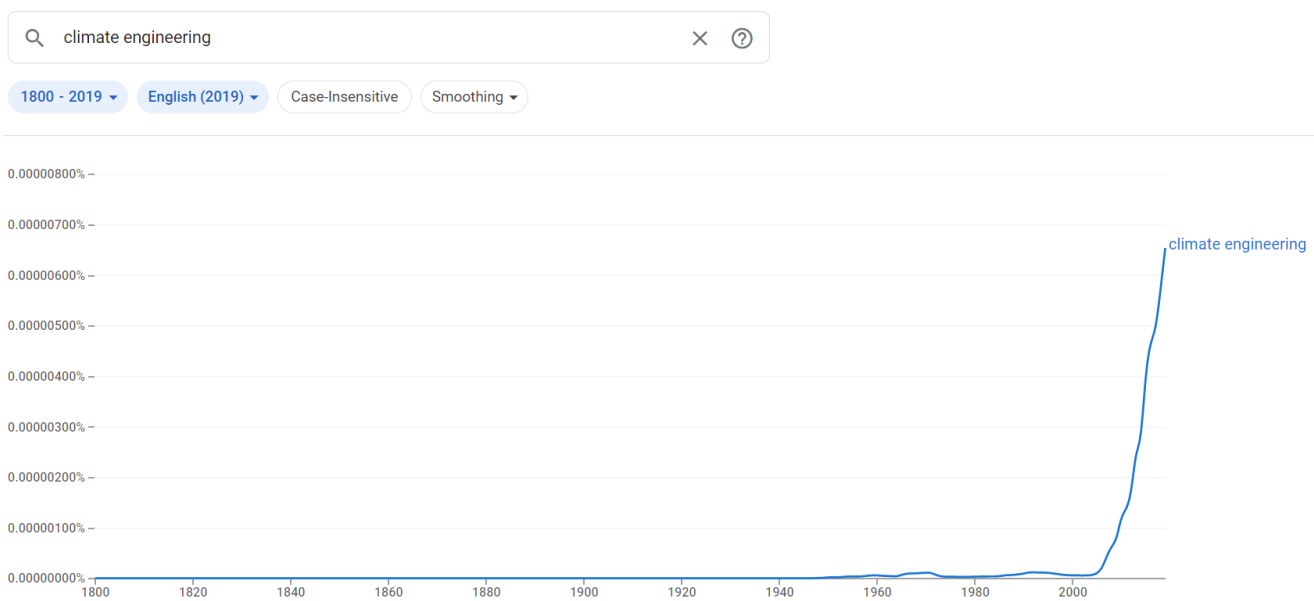## 4.1   Climate Engineering (Interaction with the planet)

Figure 3: Climate Engineering Google Ngram View Chart

4.1.1   Ethical codes

We scanned the literature searching for a set of pre-defined references to ethical codes, frameworks and guidelines within the technology family of *Climate Engineering (CE)*.

With regards to ethical *codes* within CE, the search produced six references from academic and other research organizations. Lawlor and Morely (2017) argue that given the current exacerbating climate emergency, any existent (if any) codes or ethical principles have proven insufficient. Hubert (2021) states that private entities that have an interest in engaging in CE research may be both unaware of and extricated from following the ethical standards of other professional scientists. However, in 2015, Hubert and Reichwein argue that the inter-governmental Food and Agriculture Organization (FAO) Code of Conduct for Responsible Fisheries serves as a potential exemplar for developing a code of conduct for CE research. Jinnah et al. (2019) argue that the scientific community that is conducting certain forms of CE (i.e., Solar Radiation Management (SRM)) should delimit what constitutes responsible SRM and that funders of such research should oblige researchers to comply with such a code. Along these lines, Boettcher (2019) discusses the issues with an overly broad code that may miss the mark contrasted with one that is flexible enough to be adaptable with changing technological and social needs. Morrow (2017) says that soft-law approaches to governing CE in the form of codes of conduct is such an example of a flexible and adaptive regulatory tool.

4.1.2   Ethical frameworks

When it comes to *ethical frameworks*, we found references from academia. In 2009 Morrow et al. called for the international community, including ethicists, to engage in dialogue regarding the social benefits and risks of CE research given the lack of a generally-accepted framework. Winickoff and Brown (2013) reiterate the issues regarding a clear and delimited governance framework for CE experiments and the need to clearly define them. Bellamy (2015) proposes a sociotechnical framework for CE governance that acknowledges the ethical issues of the systemic effects of the technologies of emerging sciences like CE. Svodoba (2017) mentions the United Nations Framework Convention on Climate Change (UNFCCC), Reynolds and Horton (2020) the Earth System Governance (ESG) Research Framework, and Hartzell-Nichols (2012) proposes the Precautionary Decision-Making Framework (PDMF).

4.1.3   Ethical Guidelines

Concerning *ethical guidelines*, Morrow et al. (2009) propose that ethical guidelines can be derived from the literature on ethics, specifically research involving both human and animal subjects, that can be applied directly to CE research. This is particularly so in light of CE research being relatively new and lacking clear guidelines *per se*. Hubert and Reichwein (2015) reiterate this saying that although the scientific community has general guidelines regarding research, such large-scale research in the open environment does not have any set guidelines. Further, Reynolds (2011) questions how binding such guidelines would be, saying that there are already global initiatives to develop such guidelines for CE research. Morrow (2017) argues that such guidelines for CE should be clear and qualitative and that as the scope of the impacts of research increases, so too will the proportion of the strictness of those guidelines.
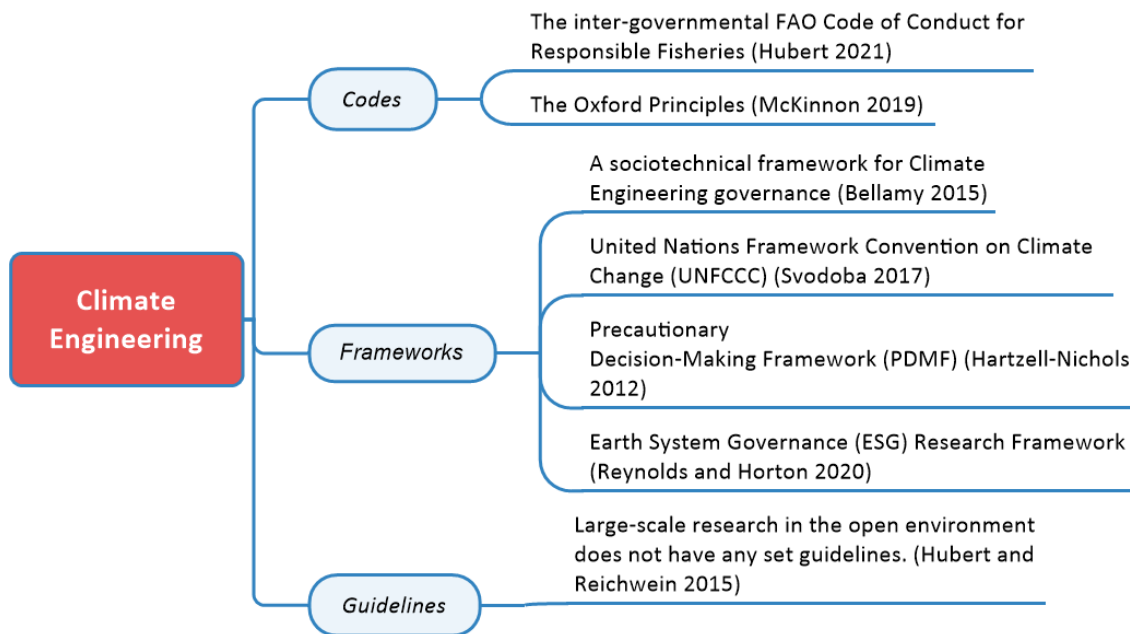
Figure 4: Mindmap of existing and published ethical codes, frameworks and guidelines for climate engineering identified in the scan. The references in bracket refer to either the original authorship or the scholar citing the frame. Proposals for new ethical codes, frameworks and guidelines for climate engineering are not included in this map.

4.1.4   Ethical principles

We then interrogated the results of the scanning further based on analytical terms suggested by Brey (2012b), see Table 3. In other words, we searched for the key ethical principles of autonomy, integrity, freedom, human rights, privacy in the documents we reviewed.

Autonomy

Discussions of autonomy are centred around citizens' rights/ability/choice to participate in CE. This may come in the form of scientists' freedom to research CE (Hubert, 2021) to citizens' ability to participate in CE research as potential subjects of its impacts. For example, both Morrow et al. (2019) and Reynolds (2011) discuss how the global impacts of CE can feasibly impact on all humans, thus raising questions of human autonomy at the global scale.

Freedom

We encountered the notion of *freedom* within two sources, one from a research organization and the other an academic document. While we understand that climate engineering raises some interesting issues related to human freedom such as the ability to live independently, the only references we found in the selected sources concerned freedom of research. We found this to be still a relevant context as it points to the dearth of research on CE and its resulting lack of information on the efficacy, risks, and benefits of geoengineering measures which will support better informed decision-making in the future. Ensuring scientific freedom in research could be an ethical measure to remedy this issue. In this regard, Hubert and Reichwein (2015) point out that freedom of research is often mentioned within guidelines for CE research, the concept is less

focused on in international law. Hubert (2021) continues by saying that although scientific freedom is often a starting point, and a central topic in CE governance approaches, it nonetheless is upheld more by the self-organizing bodies of researchers rather than international governance structures.

Human Rights

Human rights were also present in both academic documents and work from another research organization. Morrow et al. (2009) discuss how the far-reaching impacts of CE experiments require the identification of the relevant human rights that may be affected by such experiments. Hubert and Reichwein (2015) argue that the right of scientific research is often a predecessor to other human rights. As such, in the UN Declaration on Human Rights there is the right to 'share in scientific advancements and its benefits' (Reynolds, 2011). Because of this, the impacts and benefits of CE on human rights require multi-stakeholder engagement (Jinnah, 2018) and the responsibility and right for everyone to enjoy the benefits of CE research and application (Hubert, 2021).

Integrity

Integrity was an ethical issue mentioned within a number of research documents from both academic and research organisations. Hubert and Reichwein (2015) argue that there are limits to scientific freedom. There are certain obligations that scientists have in order to ethically benefit from such a freedom, this includes integrity (among other principles) in their practice of science. This notion has been better unpacked by Mitcham (2003) and linked to responsible innovation by Stilgoe et al (2013). Mitcham (2003) conceives integrity not just in science but in the science-society relationship and explains it in terms of "co-responsibility". This is made of two principles, 1) that of role responsibility, which has then undergone significant evolution from "collective responsibility" to 2) the notion of responsibility resting with a "trans-scientific community." (2003: 273).

Privacy

Privacy had only one mention in an academic source. Reynolds (2011) mentions the concerns with privacy breaches when communicating research results and the particular issues with CE research dissemination. Given the potential militarisation of CE research, the concerns of keeping potentially dual-use CE research private is of particular interest. This is particularly problematic when the experiments undertaken are not an issue, but the potential interpretation (regarding application) of the results are, and this may trigger unnecessary regulation.

4.2   Digital Extended Reality (Interaction with the digital world)

Digital Extended Reality technologies combine advanced computing systems (hardware and software) that can change how people connect with each other and their surroundings through interactions with virtual environments. Extended Reality includes artificial-intelligence-based technologies emulating or connected with human cognitive functions (e.g., voice, gesture, movement, choices, feelings), as well as human-digital machine interaction and data processing technologies to reproduce, replace, adapt, and influence human actions. [from WP1 Factsheet]

The digital extended reality technology family can be divided into enabling technologies and resulting technologies. Enabling technologies allow the resulting technologies to function and be operational. The enabling technologies may cause discrete ethical issues and be subject to separate ethical codes, frameworks and guidelines. However, the enabling technologies have been extensively studied and

analysed in other European projects and metastudies. The enabling technologies include artificial intelligence and the resulting technologies include Natural Language Processing (NLP) and Extended Reality (XR). We have organised the literature scan around the resulting technologies, since a plethora of work has already been published on the ethics of AI, etc.

Figure 5: Digital extended reality Google Ngram View Chart (note that the phrase 'Digital Extended Reality' does not have a presence and as a term pre-dates the digital era)

Figure 6: Extended reality Google Ngram View Chart

4.2.1 Codes

Only one ethical code was found related to the term 'digital extended reality'.

The reference was made about a proposed (incomplete) code of conduct for research involving VR (Madary and Metzinger 2016) which describes the Limitations of a Code of Ethics for Researchers. As they state "We would like to conclude our discussion of the research ethics of VR by noting that the proposed (incomplete) code of conduct is not intended to be sufficient for guaranteeing ethical

research in this domain. […] Scientists must understand that following a code of ethics is not the same as being ethical. A domain-specific ethics code, however consistent, developed, and fine grained of it may be, can never function as a substitute for ethical reasoning itself" (Madary and Metzinger 2016: 12).

The presence of only one reference to codes constitutes a finding. A number of explanations could apply, for example, 1) the issues are not sufficiently identified to have a complete code of conduct/ethics; 2) there is not enough consensus on the issues to solidify a code; and 3) researchers need to think critically/ethically in each particular application of digital extended reality; furthermore, 4) since ethical codes have a narrower focus than guidelines, they may well be too restrictive for the businesses that would seek to capitalise on extended digital reality and hence not yet proposed because they would be less welcome; finally 5), it is also possible that the dearth of findings in this category may be due to a methodological limitation i.e. the search term 'digital extended reality' which overlaps with too many other similar keywords to give relevant results concerning codes.

4.2.2   Ethical frameworks

Scholars justify the need to look into ethical frameworks for VR by invoking more or less explicitly, the cognate notion of acceptability. Acceptability as defined by van de Poel (2016) as a judgement that prescribes how the technology examined ought to be desirable, either instrumentally or morally (Cannizzaro et al 2020). An acceptability framework therefore is a specific part of an ethical framework. Focusing on acceptability for extended reality (XR) and using the Food and Drug Administration (FDA) Phase I-III pharmacotherapy model as guidance, Birckhead et al. (2019) states that specific phases of Virtual Reality in the context of medicine (VR clinical study designs), could undergo early testing with a focus on feasibility, acceptability, tolerability, and initial clinical efficacy. Herz and Rauschnabel (2019) argues that decision-making through VR devices requires the integration of benefits, risks, and media specific factors, all factors that pertain to ethical acceptability.

Among those who propose ethical frameworks for Extended Reality are Brey (1999), Wang and Burdon (2021) and Larson (2017). With regards to XR, Brey (1999) evokes a consequentialist framework which may be adapted to state that immoral behaviour in VR leads to harmful consequences in the physical world. Wang and Burdon (2021) mention a fundamentally ethical framework in relation to Digital Twins, virtual models of cities that are built on real-time data extracted from sensors located within a city. Wang and Burdon (2021) propose what they call a conceptual framework of trustworthiness, composed of ability, integrity and benevolence. Talking about NLP, Larson (2017) present an ethical framework for using gender as a variable in NLP, based on the scientists' commitment to expose their theoretical inclinations, their research constructs, any bias, and their methods.

4.2.3   Ethical Guidelines

A number of scholars outline the problems concerning ethical guidelines in Extended Reality and describe the existing gap in regulation (Birckhead et al 2019, Spiegel 2018, Vaidyam et al 2019, Zhou et al 2019)

Focussing on VR, Birckhead et al (2019) believes the state of current clinical VR research to be heterogeneous and cites a description of it as a "Wild West" with a lack of clear guidelines and standards. Birkhead reports concerns that current VR research is "merely descriptive" in nature, often insufficiently powered, focused on small case reports and retrospective analyses, and often does not employ experimental designs." (2019) Spiegel (2018) claims that the gap in existing guidelines is that they do not make specific recommendations for how VR manufacturers can guard the public against

the potential risks of VR and do not propose legal regulations to address the risks. With regards to NLP, Vaidyam et al (2019) note that specific guidelines on mobile health care research currently lack consensus, and Zhou et al (2019) stress the need to establish ethical guidelines for designing and implementing social chatbots to ensure that AI systems do not harm any human users.

As for citing existing guidelines, regarding NLP, Vaidyam et al (2019) mentions the World Health Organization (WHO) effort at calling for more standardized reporting outcomes for studies through mHealth Evidence Reporting and Assessment (mERA) framework.

A number of scholars and an organisation propose recommendations for ethical guidelines for extended reality. There emerge two categories of recommendations – one invoking general ethical guidelines (Slater et al 2020, UNESCO 2021), the other calling more firmly for technology-specific ethical guidelines (Guzman et al 2020, Spiegel 2018, Herz and Rauschnabel 2019). A mediating approach is proposed by Slater (2018).

At a very general level, Slater et al (2020) invoke general ethical principles to regulate the use of XR technologies and cites the United Kingdom's research ethics requirements which include respect for autonomy and dignity of persons, scientific value, social responsibility, and maximizing benefit and minimizing harm. Extending the general ethical principles to include health concerns, UNESCO (2021) invites Member States to develop guidelines for human-robot interactions and their impact on human relationships, with special attention given to the mental and physical health of human beings.

Amongst those supporting the need for tech-specific ethical guidelines, Guzman et al (2020) call for a specific set of guidelines targeted at Mixed Reality (MR), a technology that combines aspects of Augmented Reality and Virtual Reality. The ethical guidelines emphasize privileged separation among data flows in MR. Specifically, Guzman et al proposes that 1) access to sensors should elicit permission requests to the user to disaggregate access privileges, 2) there should be separate access to raw spatial data from that of the released spatial data – this can produce a privacy-preserving version of the spatial data; 3) there should be runtime access permission requests with visualizations informing users of the content the applications are desiring access to. Spiegel also proposes ethical guidelines that encompass an industry-wide rating system, legal age requirements for some VR products, informational and warning labels, public disclosure mandates, and no-share laws. Finally, along a similar line of argumentation of specific tech and industry-relevant guidelines, Herz and Rauschnabel (2019) recommends manufacturers to develop strategies to reduce risk factors with regards to users' health and privacy. One proposal would be to have third-party labels to reduce the magnitude of this risk perception within the public. With regards to privacy concerns, businesses could ensure a transparent use of data, and publish guideline restrictions for apps, and giving users more control over the way their data is used.

Slater mediates between the general approach and the specific approach to ethical guidelines by stating that in addition to the general risks of research (e.g. exposure of vulnerable people, exposure to sensitive topics, data-related issues, impact on well-being etc.), XR research must also take into account risks specific to this technology i.e. motion sickness, information overload, intensification of experience, cognitive, emotional and behavioural disturbances after re-entry into the real world following the VR experience.
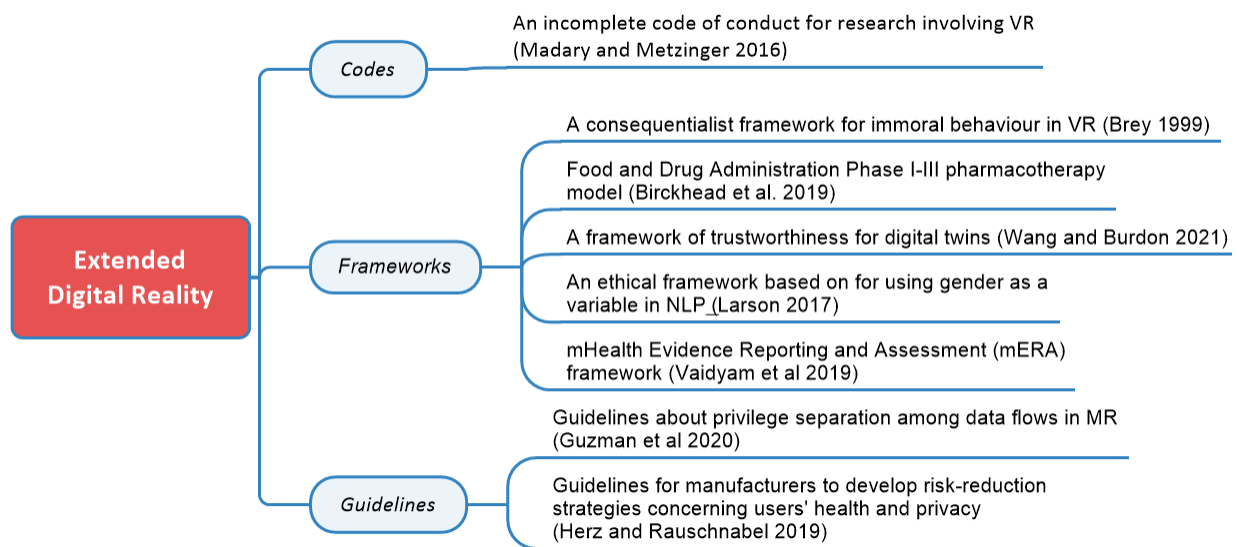
Figure 7: Mindmap of existing and published ethical codes, frameworks and guidelines for Extended Digital Reality identified in the scan. The references in bracket refer to either the original authorship or the scholar citing the frame.

4.2.4   Ethical principles

Autonomy

Spiegel explains that personal autonomy is about individual freedoms. The way in which extended reality affects autonomy and hence individual freedom, is described by O'Brolchain (2016) and Madary and Metzinger (2016). O'Brolchain (2016) outline the problems concerning autonomy arising from the convergence of Social Networks (SNs) and VR, that is 1) the threats to the knowledge condition of autonomy, 2) the threats to the freedom condition of autonomy, 3) the threats to authenticity condition of autonomy. With regards to the threats to the knowledge condition of autonomy, these consist of the filter bubble problem, the cyberbalkanization problem, the gatekeeping problem, and the distortion problem; for the threats to the freedom condition of autonomy, these consist of the addiction problem, the manipulation threat, the government threat, and the self-censorship threat; finally, concerning the threats to authenticity condition of autonomy, these are the social conformity threat, the quantified life problem, the experience machine problem, and the shallow threat. He argues that while the threats to privacy (see section below) are more known, the threats to autonomy are less-well known but equally significant. Furthermore, Madary and Metzinger (2016) explain that, if an experiment might alter a users' behaviour without their awareness of this alteration, then such an experiment could be seen as a threat to the autonomy of the subject. Another key problem is raised by Spiegel who highlights a key dilemma regarding autonomy and its subsequent effect on personal liberty. The dilemma is shown by either governments striving to maximize liberty by not regulating public use of VR, or governments striving to maximize liberty by regulating VR.

A solution to the issue of autonomy for extended reality is put forth by Madary and Metzinger (2016) who suggest that a way to preserve autonomy is simply to inform subjects of possible long-term, lasting effects.

30

Freedom

Autonomy is strictly linked to freedom. A person's autonomy relies on a sufficient degree of freedom. In relation to extended reality, and particularly to the convergence of SNs and VR, O'Brolchain (2016) outlines two key contexts that impact on freedom, that is, addiction (to SNs and VR) and surveillance from governments, as these could potentially use information gathered from these technologies to limit freedom.

Integrity and privacy

Integrity within extended reality was conceived within the umbrella issues associated with privacy. As O'Brolchain (2016) explains, physical privacy acts as a shelter against third party sensory access to an individual's body and actions and therefore concerns modesty, separateness, bodily integrity and the like. Furthermore, he explains that the threats to privacy concerning VR and SNs, include the threats to informational privacy, the continuous monitoring performed by recording devices, the unintended disclosure of information, the loss of anonymity problem, the socializing problem and the global village problem (threats to associational privacy).

As for privacy, Butz et al. (1998) underlines how important an issue it is in the design of any multi-user system within extended reality. Furthermore, and within the IoT, privacy is often associated with cognate concerns of security under the rubrics of cybersecurity. Of relevance to this discussion is the privacy part of cybersecurity, and in this regard Boeckl et al. (2019) explain how cybersecurity within IoT devices can be thought of in terms of both protecting data security and protecting individuals' privacy, both of which are high-level risk mitigation goals. Flagging up the privacy issues associated with NLP, Vaidyam (2019) notices that chatbots today often do not offer users privacy and confidentiality, for example within a clinical setting, both of which are instead assumed and protected during in-person visits to a clinician

Bender and Friedman (2018), Butz et al. (1998) and O'Brolchain (2016) make suggestions about measures that are being used to protect privacy within extended reality. Bender and Friedman (2018) explain that clear first steps would entail developing best practices for how data statements are produced, that is, finding appropriate level of detail of data, taking in consideration privacy concerns, especially for small or vulnerable populations. Within VR, Bunz assume a simple model of privacy where public objects can be seen by other users, while private objects cannot. Finally, O'Brolchain (2016) is concerned with associational privacy, where an individual would have control over excluding and including third parties in certain specific experiences.

4.3   Neurotechnologies (Interaction with the human brain - technologies and the human body)
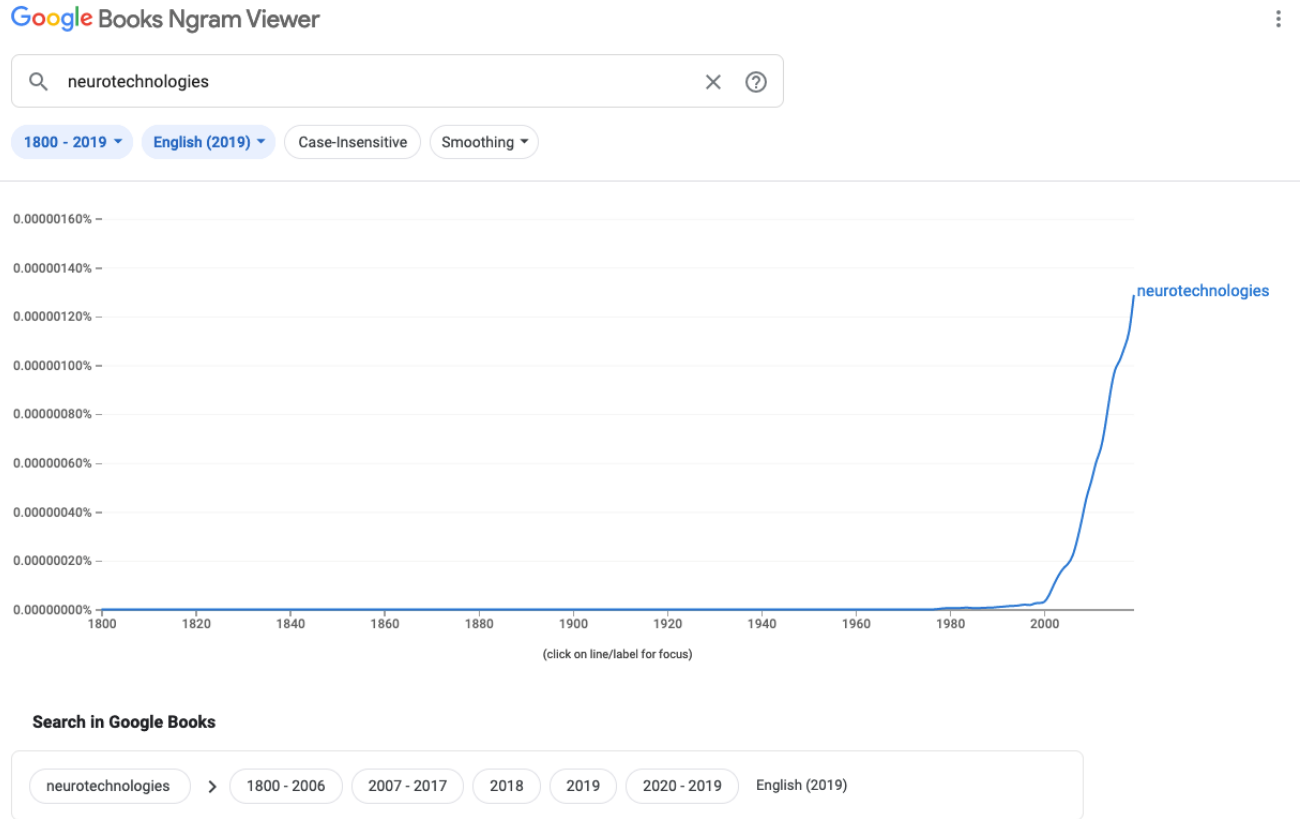
Figure 8: Neurotechnologies Google Ngram View Chart

4.3.1   Ethical codes

In terms of ethical codes of conduct within the neurotechnologies family, the search produced references from academic sources, one other research organization and one intergovernmental organization. What emerges from the literature is the diversity in the approaches used towards ethical codes of neurotechnologies – these range from an invitation towards professional self-regulation (Chang et al 2019), founding a new set of Neurorights (DSI 2020), soft-power approaches to regulation (Kreitmair 2019), cross-fertilisation with codes in separate fields (Marchant and Tournas 2019), and responsibility in technology transfer (Pfotenhauer et al 2021) and for trust (Wallach 2011).

Chang et al (2019) underlines the importance of professional self-regulation within a company, which should start by defining a code of ethics or set of clearly articulated principles to which the organisation's leadership would adhere. The Data Science Institute at Columbia University tackles ethical codes in neurotechnologies by founding a new set of codes, called the NeuroRights Initiative. This is a set of ethical codes and human rights directives that protect people from potentially harmful neurotechnologies (DSI 2020). Reflecting on regulation, particularly on a gap in governmental regulation, Kreitmar argues in favour of regulation of direct-to-consumer (DTC) neurotechnologies through what he calls a soft law approach. This consists of international codes of conducts according

to which "a group of stakeholders consisting of customers or customer representatives, neuroethicists, industry specialists, medical experts, and cybersecurity experts" (Kreitmair 2019: W2).

Relying on a cross-fertilisation strategy, akin to academic interdisciplinarity, Marchant and Tournas (2019) mention the relevance to neurotechnologies to the OECD Principles for AI (OECD 2019), the Oxford Principles to guide geoengineering (Oxford 2009), and the codes of conduct to govern nanotechnology (Bowman and Hodge 2009).

Two scholars invoke responsibility in their reflections on ethical codes. Pfotenhauer et al (2021) publish the "Nine Points to Consider", a code of good practice in university technology transfer for more responsible innovation practices, including in neurotechnologies. Wallach (2011) proposes a code made of 5 rules inviting responsible innovation, that is Rule 1: The people who design, develop or deploy a computing artefact are morally responsible for that artefact, and for the foreseeable effects of that artefact; Rule 2: a person's responsibility includes being answerable for the behaviours of the artefact and for the artefact's effects after deployment; Rule 3: People who knowingly use a particular computing artefact are morally responsible for that use. Rule 4: People who design, develop, deploy or use a computing artefact can do so responsibly only when they take into account the sociotechnical systems in which the artefact is embedded. Rule 5: People who design, develop, deploy, promote or evaluate a computing artefact should not explicitly or implicitly deceive users about the artefact's effects.

4.3.2   Ethical frameworks

When it comes to ethical frameworks, we found references from academia and other research organisations. This was the most prolific area of ethical frames as quite a lot of material revolved around ethical frameworks. As with ethical codes, there are some common approaches to ethical frameworks in neurotechnologies – underlining gaps in frameworks (HBP 2018, ABA 2019, Sadato et al 2019), the call for cross-fertilisation of ethical frameworks with those in cognate yet separate fields (Bowman et al 2018, Giordano 2014, (Ienca et al 2017) or the proposal for novel approaches and specific cases of ethical frameworks (Wingeier 2020, Goering et al 2021, Mackenzie and Walker 2015, Pfotenhauer et al. 2021).

Underlining gaps in frameworks, the Human Brain Project (HBP 2018) in Europe outlined concern with dual use concerns in relation to misuse of brain research and new computing technologies. The Australian Brain Alliance (2019) writes that Australia currently lacks a clear regulatory framework for ensuring that individuals are informed about how their data are captured, stored, analysed, and shared. Similarly, Sadato et al (2019) reports that the neuroethical framework constructed by the preceding national brain projects in Japan needed to be extended.

Calling for cross-fertilisation of ethical frameworks with those in cognate yet separate fields, Bowman et al (2018) mention two different ethical frameworks relevant to neurotechnologies – the ELSI framework that emerged from the Human Genome Project which emphasized the need for oversight structures; and the RRI framework, focussing on engaging the innovation process itself and opening the doors to more stakeholders in order to help steer technology in socially desirable directions, and argues that consistencies and differences should be considered for both frameworks for mutual learning and enrichment. Along the same line, Brindley and Giordano (2014) cite TRIPS, and ethical frameworks aiming to introduce more stringent IP rights in developing countries with greater enforcement capacity than the previous frameworks. Invoking a similar approach to framework cross-fertilisation, Ienca et al (2017) argue that neuroscience presents the issue of dual use i.e., technologies used both civil and military purposes. Hence neuroscience needs the biosecurity frameworks developed in other areas of the life sciences and that would need to involve calibrated regulation,

(neuro)ethical guidelines, and awareness-raising activities within the scientific community. The frameworks they mention include two existing U.N. treaties—the Biological Weapons Convention (BWC) and Chemical Weapons Convention (CWC)—that *de iure* should limit abuses within the neurotechnology domain; also the Universal Declaration on Bioethics and Human Rights adopted by UNESCO, but with specific focus on the challenges raised by neurotechnology. Also, they state how a neurosecurity framework could help anticipate future threats and maximize security in the neurotechnology domain.

On the contrary, Goering et al (2021) contend that cross-fertilisation of ethical frameworks may not be sufficient or appropriate. They consider the issue of commercial responsibility and regulation and note how some neurotechnologies may not fit into traditional medical regulatory frameworks and their research development may fall outside the scope of governmental regulation. To show how industry is responding to this issue, they mention IBM's effort at creating a crowdsourced, iterative framework for ethical AI called "Everyday Ethics for AI".

Amongst those who also present novel approaches to ethical frameworks for neurotechnologies are Mackenzie and Walker (2015) who propose a framework based on philosophical and ethical principles. They approach the question of whether and how neurotechnologies threaten "identity" i.e., one that is based on a relational, narrative understanding of identity and autonomy incorporating a notion of authenticity based on self-discovery and self-creation. Pfotenhauer et al. (2021) mention the case of a novel approach using the ethical principle of transparency by reporting on what they call both a framework and a code of responsibility for neurotech startup Aifred. This framework applies deep-learning algorithms to enhance individualized psychiatric treatment, which they term 'meticulous transparency' framework. They claimed this framework helped them resolve concrete design dilemmas like the use of binary predictive algorithm outputs, such as 'being' or 'not being' at risk of suicide, producing a probabilistic, rather than binary, outputs. Finally, Wingeier (2020) present the case of an 'applied' matrix-based approach, developed for the neoengineering community. They describe the IEEE Brain neuroethics framework [2] which is organized as a matrix of specific types of neurotechnologies and their current and potential applications. The specific types of neurotechnologies they mention in the framework are recording/ Sensing Technologies, medical technologies for diagnostics, stimulating/ actuating technologies, closed-loop, technologies, direct physical and biological modification technologies, augmentation and facilitation technologies; the fields of application are the medical, wellness, education, workplace, military/national security technologies, sports and competitions, entertainment, analytics, marketing & advertising, justice system technologies.

4.3.3   Ethical Guidelines

Concerning ethical guidelines, we also found references within academic, other research organisation's documents. Citing novel approaches proposed for ethical guidelines are Sadato et al (2019) and ABA (2019). DSI (2020) underlines a gap in ethical guidelines for neurotechnologies, while Kreitmair (2019 points out that it is not clear if there ought to be specific guidelines in fact. Goering and Yuste (2016) and Goering et al. (2021) support cross-fertilisation of ethical guidelines with those in cognate fields, while Kreitmair (2019) warns against a simplistic take on this approach.

Sadato et al (2019) state that the clinical research organizing team has also created guidance for modifications within ethical protocols. The Australian Brain Alliance make their goal clear: to develop

---

[2] https://brain.ieee.org/publications/ieee-neuroethics-framework/

national guidelines for responsible neuro-innovation to assist neuroscientists, engineers, and developers to translate research into effective and ethical products (ABA 2019).

Underlining the gap in guidelines by comparing the ethical challenges posed by neurotechnologies and AI, DSI (2020) outline how recent technologies based on artificial intelligence and algorithms were developed before consideration about ethics.

With regards to cross-fertilisation of guidelines, Goering and Yuste (2016) mention the Belmont report created by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research as widely respected guidelines within modern medical practice and believe that a similar set of principles should be proposed for neuroethics. Hence, they propose ethical principles and guidelines for research involving human subjects under the three core principles of respect, beneficence, and justice. Along the same line of argumentation, Goering et al. (2021) mention the European Union's ethics guidelines for AI, in which academia aligns with industry and business experts in its production.

Contrary to those who seek inspiration from the ethical regulation of medical devices, Kreitmair (2019) outlines that even those DTC (direct to consumer) neurotechnologies "that most overlap in purpose with medical devices" should not be evaluated "according to medical device criteria," because the latter have an array of safeguards in place that the former do not. Importantly, he finds that there is no stance taken on whether DTC neurotechnologies ought to be regulated. Wolpe (2002) follows a similar approach when discussing the blurring of the frontier between therapy and enhancement, a major issue with neurotechnologies. Discussing the work of bioethicist Norm Daniels (1985, cited in Wolpe 2002: 389) he points at the difficulty of creating what he calls a meaningful – in lieu of ethics – enhancement standard to use to allocate medical care or create guidelines for clinical treatment.
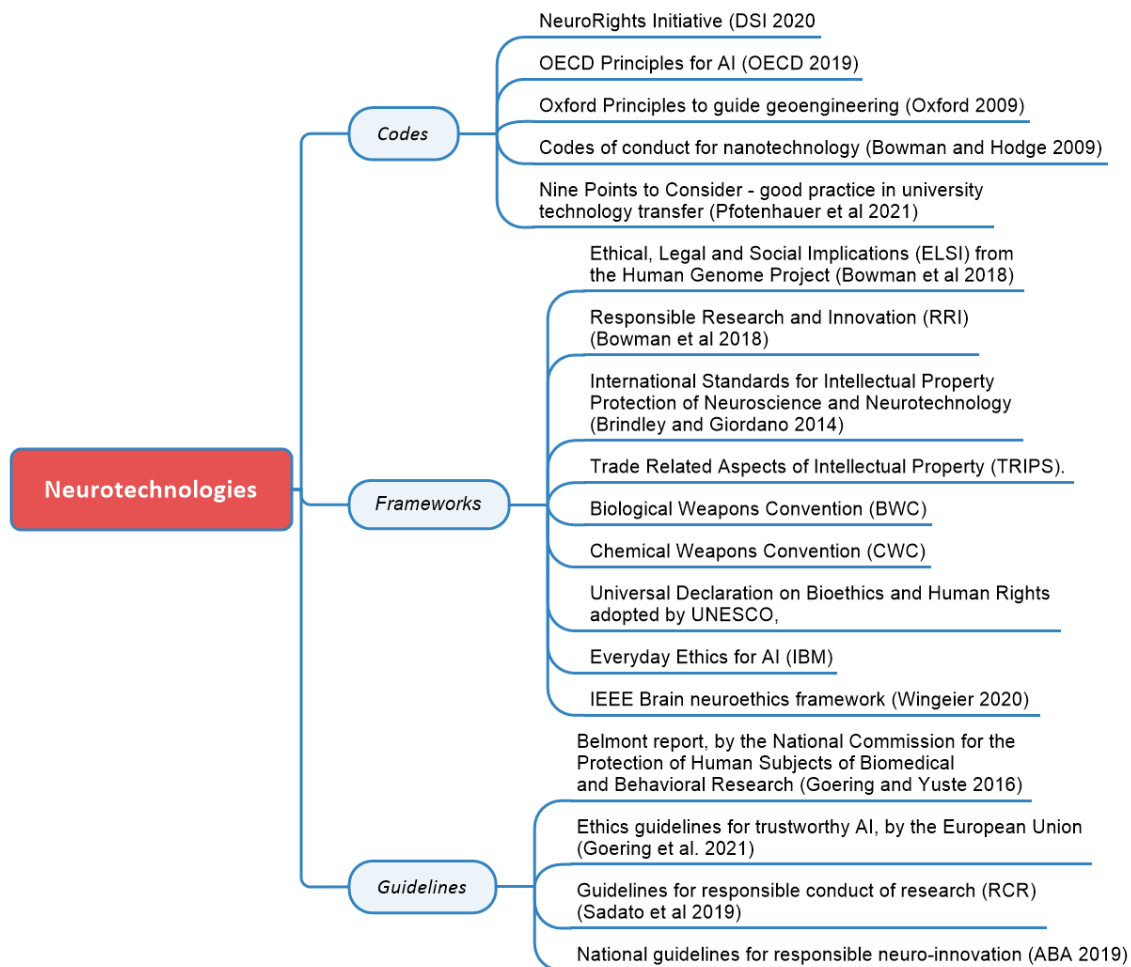
Figure 9: Mindmap of existing and published ethical codes, frameworks and guidelines for neurotechnologies identified in the scan. The references in bracket refer to either the original authorship or the scholar citing the frame. Proposals for new ethical codes, frameworks and guidelines for neurotechnologies are not included in this map.

4.3.4   Ethical principles

We also scanned the same literature to gain a preliminary idea of how documents foregrounding ethical codes, frameworks and guidelines, contextualise issues such as autonomy, freedom, human rights, integrity and privacy within these ethical frames.

Autonomy

We came across references to autonomy in academic documents and intergovernmental organisations. Mainly these dealt with examples of violation of autonomy when using neurotechnologies in general, but also with uncertainty surrounding specific examples of neurotechnologies.

OECD (2021) recognise the centrality of autonomy for neurotechnologies and the fact that these technologies can raise ethical, legal, and societal questions based on this principle. They define autonomy as the freedom to make one's own choices.

A number of scholars mention the potential to compromise autonomy as an ethical issue of neurotechologies. Referring to neurotechnologies in general, Farah (2015) states that neurotechnologies pose a risk as it would be a violation of a subject's right to autonomyto be enrolled as a research subject without informed consent, and neurotechnologies would favour that. For Mackenzie and Walker (2015) autonomy violation could happen in relation to the question of whether and how neurotechnologies threaten identity. Pfotenhauer et al. 2021 also underline the importance of avoiding autonomy violation, particularly in relation to becoming vulnerable to manipulation through use of neurotechnologies. They explain how many emerging app-based neurotechnologies that aim to perform mental health interventions collect data to predict or respond to cognitive states. However, at the same time this can make individuals vulnerable to manipulation for surveillance, policing, and economic or political reasons.

Along the same line, but making a reference to specific examples of neurotechnologies, Sadako et al (2019) are uncertain about how fMRI neurofeedback could protect patient autonomy, despite its usefulness for innovatively treating patients with neuropsychiatric disorders. Wallach (2011) mentions the uncertainty for autonomy of another case of a specific neurotechnology. He states there is no understanding of whether combining various cognitive enhancers with neuroprosthetics will optimise the freedom of individuals or undermine their autonomy.

Freedom

We encountered the notion of freedom within academic documents. Ienca and Andorno (2017) describe how neurotechnologies can cause ethical issues at the level of freedom – they state the technologies have the potential to allow access to aspects of mental information. Along a similar line, Goering et al. (2021) argue for a conceptual re-thinking of already recognized rights such as the freedom of thought. Ienca and Andorno (2017) term the freedom related to the use of neurotechnologies as 'cognitive liberty, which for them, resonates with the more common notion of 'freedom of thought'. Citing Sententia (2004), Ienca and Andorno (2017) also state that cognitive liberty, or the freedom to control one's own consciousness is the underlying element of every other type of freedom.

Human rights

Concerns for human rights in neurotechnologies were present in both academic documents and in one document produced by intergovernmental organisation.

Farah (2015) provides a definition of human rights, I.e. rights are moral entitlements, or ''must-haves''. Several scholars and one organisation spell out the relevance of human rights for neurotechnologies. For example, OECD (2021) emphasises the centrality of human rights in order to avoid harm in the use of neurotechnologies. Brindley and Giordano (2014) outline a gap in the consideration of human rights within this emergent technology family, particularly by underlining how the WTO places economic considerations ahead of concerns for human rights. In regard to the need to consider human rights within neurotechnologies, Goering and Yuste (2016) argue that given the use of methods that may substantially alter one's personality in neurotechnologies, what they name as 'private internal spaces' and 'agential identity' need to be integrated into our understandings of human rights. As testament to this effort, Ienca et al (2017) mention a 2010 pledge by neuroscientists in 17 different countries. Their concern was with dual use of neurotechnologies and they proposed two obligations underlining 1) an awareness of potential applications especially those that violate basic human rights as in the case of the technology being used for torture and aggressive war'' 2) a refusal to participate knowingly in the application of neuroscience to violations of basic human rights (Bell, 2014 cited in Ienca et al 2017).

A few scholars name specific rights being relevant to neurotechnologies. Goering et al. (2021) do so by mentioning the example of specific neurotechnologies. According to them, brain computer interfaces (BCI) and deep brain stimulators (DBS), will have profound implications for society and human rights, and recommend the establishment of new "Neurorights". They name these as mental liberty, mental privacy and mental integrity. Along the same line of naming the new human rights. Also Ienca and Andorno (2017) name the new neurotechnology-relevant human rights that is, the right to cognitive liberty, the right to mental privacy, the right to mental integrity, and the right to psychological continuity.

There appears to be more concern for new human rights in neurotechnologies than proposals for it, so far. The only example found include what DSI (2020) describe as the NeuroRights Initiative, puts forth human rights directives that protect people from potential harm caused by neurotechnologies.

Integrity

Integrity was an ethical issue mentioned only within academic source and in one document produced by intergovernmental organisation.

Ienca and Andorno (2017) acknowledge the need to consider the importance of mental integrity in neurotechnologies, for example. That is because they believe the artificial alteration of a person's neural processes by means of technologies pose an unprecedented threat to that person's mental integrity. Similarly, in their recommendations for responsible development of neurotechnologies, Goering et al. (2021) also underline the importance of mental integrity amongst the new Neurorights. They explain how these regulations serve to protect bodily integrity and avoid exploitation of brain data for commercial purposes, by eliminating monetary incentives. In this respect, Mackenzie and Walker (2015) pose a fundamental question (citing Glannon 2008) about how much intervention by mean of neurotechnologies can one's life accommodate without threatening the integrity of the whole, and therefore radically altering the identity of the person.

As a way forward to tackle the ethical right to integrity in neurotechnology, OECD (2021) call for an intervention across the sector of neurotechnology, according to which stakeholders should pursue the development of best practices and consider several ethical issues when conducting business, including accountability, transparency, integrity, trustworthiness, responsiveness, and safety. Another proposal that appears to be more underway, is that recalled by Pfotenhauer et al. (2021). They mention the case of the Chilean senate which is considering a constitutional amendment to legally codify 'neurorights' to protect the mental integrity its citizens.

Privacy

Privacy in neurotechnologies was the most popular ethical issue being raised, across both academic as well as other research organisations' documents, and spanning countries of production such as Australia, Japan, and Switzerland. Reflections on privacy issues in neurotechnologies revolved around definitions of privacy (Ienca and Andorno 2017, Goering and Yuste 2016), outlining the problems connected with privacy (Wallach 2011, Pfotenhauer et al. (2021), outlining existing gaps in both regulation and industry in dealing with neurotechnologies (Sadato et al 2019, OECD 2021) and propose areas of intervention (ABA 2019, Ienca and Andorno, 2017).

In regard with definitions of privacy, Ienca and Andorno (2017) state that privacy within neurotechnologies includes not only the issue access to personal information, but also that of access to bodies and private places. Similarly, Goering and Yuste (2016) explain that privacy in neurotechnologies will need to take into account the privacy of our internal lives (citing Farah and Wolpe, 2004).

Scholars circumscribe the set of problems concerning privacy in neurotechnologies: Goering et al. (2021) underline how privacy is important for this emerging technology not just because information can be 'scraped' or mined from brain data but because neurotechnologies allow for new ways of "writing" information into the brain. Wallach (2011) very practically warns about how introducing robots into the home and other social settings raises privacy risks similar to those posed by surveillance cameras. Pfotenhauer et al. (2021) discuss brain data privacy in relation to industry. Businesses developing neurotechnologies will have to deal with data ownership, security, privacy and consent. Furthermore, they argue that at brain data privacy within neurotechnologies involves potentially more sensitive data than dealt with in other domains of technology, particularly as the data harvested through neurotechnologies could be manipulated for legal reasons and used in court.

Amongst those who outlining existing gaps in knowledge of privacy issue in both regulation and industry, Ienca and Andorno (2017) declare that there is no specific legal or technical safeguard that protects brain data from being harvested as other types of information. Sadato et al (2019) predict that if de-identification of data was transparent within neurotechnologies, Japanese citizens concerned about privacy of brain data, would feel more secure (Sadato et al 2019). Transposing the issue onto the industry, OECD (2021) warn that potential business models will have to address (brain) data privacy when trading with neurotechnologies.

A few scholars put forth suggestions concerning the safeguard of brain data. The ABA (2019) underline that it is important to address how personal information contained in brain data is shared with third parties. (Ienca and Andorno, 2017) call for the updating privacy rights to account for mental privacy, and DSI (2020) name a proposal for doing so, that is, The Right to Mental Privacy. This refers to the idea that data obtained from scrutinizing neural activity should be kept private, and the sale of such neural data should be strictly regulated.

5.Conclusion and grounds for further work

This report is based on Deliverable 2.1 and comprises of two parts, firstly, a brief review of three approaches to ethical analysis, ATE, eTA and Future Studies and secondly, the scanning of existing ethical codes, guidelines and frameworks in the ethical literature related to the three technology families of climate engineering, digital extended reality and neurotechnologies.

In the first part of this work, the brief review of three approaches to ethical analysis, ATE, eTA and Future Studies show that it is difficult to predict the future, particularly considering that the very notion of future is a non-neutral, politically-charged concept that ought to be subjected to critical analysis and scrutinised itself for any bias – be it gender or economics-driven. However, as techniques and approaches they each demonstrate that it is possible to develop some guidance on how to assess the possible ethical issues associated with a specific technology, so that developers and users may reflect on this and potentially incorporate those reflections into their design, development and use. Through the review of ATE, eTA and Future Studies we have prepared the ground for analysis by circumscribing a framework for the ethical analysis which will be further developed in the following tasks of WP2. The framework is that proposed by Brey (2012a) under the rubrics of ATE, proposing three levels of ethical analysis -technology, artifact and application level – and defining the 'objects of ethical analysis' for each of these levels, as properties or processes that might lead to ethical issues. As an outcome of this review, we will look to draw further on the other two approaches, to incorporate their insights into the ethical analysis of the selected technology families. Given the importance of these issues, TechEthos proposes to further develop this broad approach to ethical analysis in Task 2.2.

Furthermore, in the second part of this work, that is, the scanning of the literature related to the three technology families - we unpacked several key ethical issues – some of which were specific to the emerging technology, others cut across the three domains. Ethical paradigms are cultivated primarily in the Western juridical-legal-ethical context, and because of this issues to do with personal autonomy, freedom, integrity, human rights and privacy were highlighted in the analysis of the texts. When searching for these ethical principles within the selected documents, we often came across other themes mentioned within the same paragraph or group of paragraphs, which were covered in our list. Although we did not specifically pre-select the themes of dignity and trust, it was surprising that such themes were less obvious in association with discussions of autonomy, freedom, integrity, human rights and privacy in the selected literature.

We identified common research trends when scanning the experts' views on ethics of emerging technologies: 1) researchers would often outline the gaps in existing regulations, 2) they would then mention specific existing regulatory codes, frameworks or guidelines, and 3) they would either advocate for cross-fertilisation of existing areas (for example the life sciences and legal scholars and practitioners) or advocate strongly against them in favour of specialisation and expertise in specific areas, and finally, 4) propose novel approaches and present case studies of application of ethical codes, guidelines and frameworks to tackle the ethics of the emerging technologies.

Following the European Union experience of the emergence of Artificial Intelligence (AI) and its impact on society, there was concern that ethics lagged behind technological innovation for emerging technologies with high socio-economic impact. Ensuring that innovation and ethics co-develop is the primary motivation behind the TechEthos project. The hope is that the outcome of the analysis will allow us to feed into the development process of the technology families, while their trajectory can still be influenced.

This is even more pressing, as we frequently found academics took a 'wild west' approach and issued guidelines that sometimes contradicted or did not take into account existing laws and codes but innovated their own based on their particular research areas and expertise. Unlike a law, which requires the commitment of elected representatives, codes, frameworks and guidelines are produced in both academic and grey literature. Researchers in responsible research and innovation (RRI) are actively engaged in promoting and advocating for specific legal changes. This can create a representational bias (expert-driven) that is shaping both the technology as well as the ethical responses to it.

A number of these technological innovations are forcing us to reconsider some of our core ethical values and this can be explored further in future work in this WP and others.

6.References

Accenture (2021) Responsible AI: From principles to practice, Available from
https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice

Bellamy, Rob. "A Sociotechnical Framework for Governing Climate Engineering." Science, Technology, & Human Values 41, no. 2 (March 1, 2016): 135–62. https://doi.org/10.1177/0162243915591855

Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041

Benton, A., Coppersmith, G., & Dredze, M. (2017). Ethical Research Protocols for Social Media Health Research. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 94–102. https://doi.org/10.18653/v1/W17-1612

Birckhead, B., Khalil, C., Liu, X., Conovitz, S., Rizzo, A., Danovitch, I., Bullock, K., & Spiegel, B. (2019). Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. *JMIR Mental Health*, *6*(1), e11973. https://doi.org/10.2196/11973

Boeckl, K., Fagan, M., Fisher, W., Lefkovitz, N., Megas, K. N., Nadeau, E., O'Rourke, D. G., Piccarreta, B., & Scarfone, K. (2019). Consi*derations for managing Internet of Things (IoT) cybersecurity and privacy risks* (NIST IR 8228; p. NIST IR 8228). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8228

Boettcher, Miranda. "Cracking the Code: How Discursive Structures Shape Climate Engineering Research Governance." Environmental Politics 29, no. 5 (July 28, 2020): 890–916. https://doi.org/10.1080/09644016.2019.1670987.

de Boer, B., Hoek, J. and Kudina, O., 2018. Can the technological mediation approach improve technology assessment? A critical view from 'within'. *Journal of responsible innovation*, *5*(3), pp.299-315.

Bowman, D.M. *et al.* (2018) 'The neurotechnology and society interface: responsible innovation in an international context', *Journal of Responsible Innovation*, 5(1), pp. 1–12. doi:10.1080/23299460.2018.1433928.

Brey, P. (1999) 'The ethics of representation and action in virtual reality', *Ethics and Information Technology*, 1(1), pp. 5–14. doi:10.1023/A:1010069907461.

Brey, P.A.E. (2012a) 'Anticipating ethical issues in emerging IT', *Ethics and Information Technology*, 14(4), pp. 305–317. doi:10.1007/s10676-012-9293-y.

Brey, P.A.E. (2012b) 'Anticipatory Ethics for Emerging Technologies', *NanoEthics*, 6(1), pp. 1–13. doi:10.1007/s11569-012-0141-7.

Bringsjord, S., Bello, P., & Ferrucci, D. (2001). Creativity, the Turing Test, and the (Better) Lovelace Test. *Minds and Machines*, *11*(1), 3–27. https://doi.org/10.1023/A:1011206622741

Brindley, T. and Giordano, D.J. (2014) 'International Standards for Intellectual Property Protection of Neuroscience and Neurotechnology: Neuroethical Legal and Social (NELS) Considerations in

Light of Globalization', Available at: https://www-cdn.law.stanford.edu/wp-content/uploads/2018/05/brindley.pdf (Accessed: 2 August 2021).

O'Brolcháin, F., Jacquemard, T., Monaghan, D., O'Connor, N., Novitzky, P., & Gordijn, B. (2016). The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy. *Science and Engineering Ethics*, *22*(1), 1–29. https://doi.org/10.1007/s11948-014-9621-1

Butz, A., Beshers, C., & Feiner, S. (1998). Of Vampire mirrors and privacy lamps: privacy management in multi-user augmented environments. *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology - UIST '98*, 171–172. https://doi.org/10.1145/288392.288598

Cannizzaro S, Procter R, Ma S, Maple C (2020) Trust in the smart home: Findings from a nationally representative survey in the UK. PLoS ONE 15(5): e0231615. https://doi.org/10.1371/journal.pone.0231615

Criado, C.P. (2019). Invisible Women: Exposing Data Bias in a World Designed by Men. London, Random House.

CHALMERS, D. (2022). *REALITY +: a philosophical journey through virtual worlds*. ALLEN LANE.

Chalmers, D. J. (2017). The Virtual and the Real. *Disputatio*, *9*(46), 309–352. https://doi.org/10.1515/disp-2017-0009

De Brigard, F. (2010). If you like it, does it matter if it's real? *Philosophical Psychology*, *23*(1), 43–57. https://doi.org/10.1080/09515080903532290

De Guzman, J. A., Thilakarathna, K., & Seneviratne, A. (2020). Security and Privacy Approaches in Mixed Reality: A Literature Survey. *ACM Computing Surveys*, *52*(6), 1–37. https://doi.org/10.1145/3359626

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & MÃ¼hlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00026

DSI (2020) 'Researchers create guidelines to ensure ethical development of neurotechnology' MedicalXpress. Available from https://medicalxpress.com/news/2020-03-guidelines-ethical-neurotechnology.html (Accessed: 2 August 2021).

Farah, M.J. (2015) 'An Ethics Toolbox for Neurotechnology', *Neuron*, 86(1), pp. 34–37. doi:10.1016/j.neuron.2015.03.038.

Feiner, S. K. (1999). The importance of being mobile: some social consequences of wearable augmented reality systems. *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, 145–148. https://doi.org/10.1109/IWAR.1999.803815

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Franks, M. A. (2011). Unwilling Avatars: Idealism and Discrimination in Cyberspace. *Columbia Journal of Gender and Law*, Vol. 20 No. 2 (2011). https://doi.org/10.7916/CJGL.V20I2.2621

Giordano, J., Forsythe, C. and Olds, J. (2010) 'Neuroscience, Neurotechnology, and National Security: The Need for Preparedness and an Ethics of Responsible Action', *AJOB Neuroscience*, 1(2), pp. 35–36. doi:10.1080/21507741003699397.

Goering, S. *et al.* (2021) 'Recommendations for Responsible Development and Application of Neurotechnologies', *Neuroethics* [Preprint]. doi:10.1007/s12152-021-09468-6.

Goering, S. and Yuste, R. (2016) 'On the Necessity of Ethical Guidelines for Novel Neurotechnologies', *Cell*, 167(4), pp. 882–885. doi:10.1016/j.cell.2016.10.029.

Greylit (2021) What is Grey Literature? Available from https://www.greylit.org/about

Gurzawska, A. (2021) 'Responsible Innovation in Business: Perceptions, Evaluation Practices and Lessons Learnt', *Sustainability,* 13(4), p. 1826. doi:10.3390/su13041826.

Hartzell-Nichols, L. (2012) 'Precaution and Solar Radiation Management', *Ethics, Policy & Environment*, 15(2), pp. 158–171. doi:10.1080/21550085.2012.685561.

HBP (2018) Opinion on Responsible Dual Use from the Human Brain Project. Available from https://www.humanbrainproject.eu/en/follow-hbp/news/opinion-on-responsible-dual-use-from-the-human-brain-project/

Hester, K. *et al.* (2015) 'Anticipatory Ethics and Governance (AEG): Towards a Future Care Orientation Around Nanotechnology', *NanoEthics*, 9(2), pp. 123–136. doi:10.1007/s11569-015-0229-y.

Herz, M., & Rauschnabel, P. A. (2019). Understanding the diffusion of virtual reality glasses: The role of media, fashion and technology. *Technological Forecasting and Social Change*, *138*, 228–242. https://doi.org/10.1016/j.techfore.2018.09.008

Hubert, Anna-Maria. "A Code of Conduct for Responsible Geoengineering Research." Global Policy 12, no. S1 (2021): 82–96. https://doi.org/10.1111/1758-5899.12845.

Hubert, Anna-Maria, and David Reichwein. "An Exploration of a Code of Conduct for Responsible Scientific Research Involving Geoengineering," n.d., 96.

Jinnah, Sikina. "Why Govern Climate Engineering? A Preliminary Framework for Demand-Based Governance." International Studies Review 20, no. 2 (June 1, 2018): 272–82. https://doi.org/10.1093/isr/viy022.

Jinnah, Sikina, Simon Nicholson, David R. Morrow, Zachary Dove, Paul Wapner, Walter Valdivia, Leslie Paul Thiele, et al. "Governing Climate Engineering: A Proposal for Immediate Governance of Solar Radiation Management." Sustainability 11, no. 14 (July 20, 2019): 3954. https://doi.org/10.3390/su11143954.

Kiran, A.H., Oudshoorn, N. and Verbeek, P.P., 2015. Beyond checklists: toward an ethical-constructive technology assessment. *Journal of responsible innovation*, 2(1), pp.5-19.

Knobl, W. 2019. "Contingency" In P In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp.53-60/

Kreitmair, K.V. (2019) 'Response to Open Peer Commentaries on "Ethical Dimensions of Direct-to-Consumer Neurotechnologies"', *AJOB Neuroscience*, 10(4), pp. W1–W3. doi:10.1080/21507740.2019.1675798.

La Fors, K., Custers, B. and Keymolen, E. (2019) 'Reassessing values for emerging big data technologies: integrating design-based and application-based approaches', *Ethics and Information Technology*, 21(3), pp. 209–226. doi:10.1007/s10676-019-09503-4.

Larson, B. (2017). Gender as a Variable in Natural-Language Processing: Ethical Considerations. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 1–11. https://doi.org/10.18653/v1/W17-1601

Lawlor, R., and H. Morley. "Climate Change and Professional Responsibility: A Declaration of Helsinki for Engineers." Science and Engineering Ethics 23, no. 5 (October 2017): 1431–52.

Le Quin, U. 2016 [1969] Introduction. The Left Hand of Darkness. Narrated by George Guidall. Audo book. Audible Recorded Books.

Liberati, N. (2016). Augmented reality and ubiquitous computing: the hidden potentialities of augmented reality. *AI & SOCIETY*, *31*(1), 17–28. https://doi.org/10.1007/s00146-014-0543-x

Lombard, M., & Ditton, T. (2006). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), 0–0. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Madary, M., & Metzinger, T. K. (2016). Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI*, *3*. https://doi.org/10.3389/frobt.2016.00003

'Mackenzie and Walker 2015' (no date). Available at: https://link.springer.com/content/pdf/10.1007%252F978-94-007-4707-4_10.pdf (Accessed: 2 August 2021).

Mackenzie, C. and Walker, M. (2015) 'Neurotechnologies, Personal Identity, and the Ethics of Authenticity', in Clausen, J. and Levy, N. (eds) *Handbook of Neuroethics*. Dordrecht: Springer Netherlands, pp. 373–392. doi:10.1007/978-94-007-4707-4_10.

Madary, M. and Metzinger, T.K. (2016) 'Recommendations for Good Scientific Practice and the Consumers of VR-Technology', *Frontiers in Robotics and AI*, 3. doi:10.3389/frobt.2016.00003.

Marak, S. 2019. "Anthropocene' In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp 19-24.

Marchant, G. and Tournas, L. (2019) 'Filling the Governance Gap: International Principles for Responsible Development of Neurotechnologies', *AJOB Neuroscience*, 10(4), pp. 176–178. doi:10.1080/21507740.2019.1665135.

Markus, M.L. and Mentzer, K. (2014) 'Foresight for a responsible future with ICT', *Information Systems Frontiers*, 16(3), pp. 353–368. doi:10.1007/s10796-013-9479-9.

Sundvall, S. 2019. "Artificial Intelligence" In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp 29-34.

Paul. H. 2019. "Introduction" In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp 1-4.

McKinnon, C. (2019) 'Sleepwalking into lock-in? Avoiding wrongs to future people in the governance of solar radiation management research', *Environmental Politics*, 28(3), pp. 441–459. doi:10.1080/09644016.2018.1450344.

Mitcham, C. (2003) 'Co-responsibility for research integrity', *Science and Engineering Ethics*, 9(2), pp. 273–290. doi:10.1007/s11948-003-0014-0.

Mooradian, N. (2006). VIRTUAL REALITY, ONTOLOGY, AND VALUE. *Metaphilosophy*, *37*(5), 673–690. https://doi.org/10.1111/j.1467-9973.2006.00460.x

Morrow, David. "International Governance of Climate Engineering: A Survey of Reports on Climate Engineering, 2009-2015." SSRN Electronic Journal, 2017. https://doi.org/10.2139/ssrn.2982392.

Morrow, David R, Robert E Kopp, and Michael Oppenheimer. "Toward Ethical Norms and Institutions for Climate Engineering Research." Environmental Research Letters 4, no. 4 (October 2009): 045106. https://doi.org/10.1088/1748-9326/4/4/045106.

Munoko, I., Brown-Liburd, H.L. and Vasarhelyi, M. (2020) 'The Ethical Implications of Using Artificial Intelligence in Auditing', *Journal of Business Ethics*, 167(2), pp. 209–234. doi:10.1007/s10551-019-04407-1.

Müller, O. and Rotter, S. (2017) 'Neurotechnology: Current Developments and Ethical Issues', *Frontiers in Systems Neuroscience*, 0. doi:10.3389/fnsys.2017.00093.

OECD Legal Instruments (no date). Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0457 (Accessed: 2 August 2021).

Palmas, F., & Klinker, G. (2020). Defining Extended Reality Training: A Long-Term Definition for All Industries. *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, 322–324. https://doi.org/10.1109/ICALT49669.2020.00103

Palm, E. and Hansson, S.O., 2006. The case for ethical technology assessment (eTA). *Technological forecasting and social change*, *73*(5), pp.543-558.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Potts, J., 2018. Futurism, futurology, future shock, climate change: Visions of the future from 1909 to the present. *Portal: Journal of Multidisciplinary International Studies*, *15*(1-2), pp.99-116.

Pfotenhauer, S.M. *et al.* (2021) 'Mobilizing the private sector for responsible innovation in neurotechnology', *Nature Biotechnology*, 39(6), pp. 661–664. doi:10.1038/s41587-021-00947-y.

Ramirez, E. J. (2019). Ecological and ethical issues in virtual reality research: A call for increased scrutiny. *Philosophical Psychology*, *32*(2), 211–233. https://doi.org/10.1080/09515089.2018.1532073

Reijers, W. et al. (2018) 'Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations', *Science and Engineering Ethics*, 24(5), pp. 1437–1481. doi:10.1007/s11948-017-9961-8.

Reynolds, J.L. and Horton, J.B., (2020) An earth system governance perspective on solar geoengineering. Earth System Governance, 3, p.100043.

Reynolds, Jesse. "The Regulation of Climate Engineering." Law, Innovation and Technology 3, no. 1 (July 2011): 113–36. https://doi.org/10.5235/175799611796399821.Roesner, F., Denning, T.,

Newell, B. C., Kohno, T., & Calo, R. (2014). Augmented reality: hard problems of law and policy. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 1283–1288. https://doi.org/10.1145/2638728.2641709

Ross, D.G., Oppegaard, B. and Willerton, R. (2019) 'Principles of Place: Developing a Place-Based Ethic for Discussing, Debating, and Anticipating Technical Communication Concerns', *IEEE Transactions on Professional Communication*, 62(1), pp. 4–26. doi:10.1109/TPC.2018.2867179.

Sadato, N. *et al.* (2019) 'Neuroethical Issues of the Brain/MINDS Project of Japan', *Neuron*, 101(3), pp. 385–389. doi:10.1016/j.neuron.2019.01.006.

Sardar, Z., (2010) The Namesake: Futures; futures studies; futurology; futuristic; foresight—What's in a name?. *Futures*, *42*(3), pp.177-184.

Shilton, K. (2015) 'Anticipatory Ethics for a Future Internet: Analyzing Values During the Design of an Internet Infrastructure', *Science and Engineering Ethics*, 21(1), pp. 1–18. doi:10.1007/s11948-013-9510-z.

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D., & Silver, J. (2020) The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality*, *1*, 1. https://doi.org/10.3389/frvir.2020.00001

Spiegel, J. S. (2018). The Ethics of Virtual Reality Technology: Social Hazards and Public Policy Recommendations. *Science and Engineering Ethics*, *24*(5), 1537–1550. https://doi.org/10.1007/s11948-017-9979-y

Spengler. B. 2019. "Imagination" In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp 163-170.

Schneider. R. 2019 "Gesture" In Paul, H. ed., 2019. *Critical terms in futures studies*. Palgrave-MacMillan, Switzerland. Pp 145-150.

Stahl, B. (2011). IT for a better future: how to integrate ethics, politics and innovation. *Journal of Information, Communication & Ethics in Society*, 9(3), 140–156.

Stilgoe, J., Owen, R. and Macnaghten, P., (2013) Developing a framework for responsible innovation. *Research policy*, 42(9), pp.1568-1580.

Svoboda, T. (2017) *The Ethics of Climate Engineering: Solar Radiation Management and Non-Ideal Justice*. 1st edn. Routledge. doi:10.4324/9781315468532.

UNESCO. (2021). *Draft text of the Recommendation on the Ethics of Artificial Intelligence*. Intergovernmental Meeting of Experts (Category II) related to a Draft Recommendation on the Ethics of Artificial Intelligence, online, 2021, Paris. https://unesdoc.unesco.org/ark:/48223/pf0000377897

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Wallach, W. (2011) 'From Robots to Techno Sapiens: Ethics, Law and Public Policy in the Development of Robotics and Neurotechnologies', *Law, Innovation and Technology*, 3(2), pp. 185–207. doi:10.5235/175799611798204888.

Wang, B.T. and Burdon, M. (2021) 'Automating Trustworthiness in Digital Twins', in Wang, B.T. and Wang, C.M. (eds) *Automating Cities*. Singapore: Springer Singapore (Advances in 21st Century Human Settlements), pp. 345–365. doi:10.1007/978-981-15-8670-5_14.

Weckert, J. (2002). Lilliputian Computer Ethics. *Metaphilosophy*, *33*(3), 366–375. https://doi.org/10.1111/1467-9973.00232

Wingeier, B. (2020) 'Get Engaged » Help create the guidelines in your area of expertise', Available at: http://brain.ieee.org/wp-content/uploads/sites/52/2020/03/ieee_brain_neuroethics_framework_double-sided-031920.pdf (Accessed: 02 August 2021).

Wright, D. (2011) 'A framework for the ethical impact assessment of information technology', *Ethics and Information Technology*, 13(3), pp. 199–226.

Woolgar, S. (Ed.). (2002). *Virtual society? technology, cyberbole, reality*. Oxford University Press.

Yuste, R. *et al.* (2017) 'Four ethical priorities for neurotechnologies and AI', *Nature*, 551(7679), pp. 159–163. doi:10.1038/551159a.

Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2019). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *ArXiv:1812.08989 [Cs]*. http://arxiv.org/abs/1812.08989

TECHETHOS

FUTURE ○ TECHNOLOGY ○ ETHICS

Coordinated by

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

Partners

Airi ASSOCIAZIONE ITALIANA PER LA RICERCA INDUSTRIALE

allea | All European Academies

cea

dmu.ac.uk DE MONTFORT UNIVERSITY LEICESTER

ecsite EUROPEAN NETWORK SCIENCE CENTRES & MUSEUMS

eurec

TRILATERAL RESEARCH

TUDelft

UNIVERSITY OF TWENTE.

Linked Third Parties

BUCHAREST SCIENCE FESTIVAL

CENTER FOR THE PROMOTION OF SCIENCE

iQ LANDIA

Consorcio Parque de las Ciencias

ScienceCenter NETZWERK

Vetenskap & Allmänhet VA – PUBLIC & SCIENCE

🐦 @TechEthosEU   in TechEthosEU

www.techethos.eu   info@techethos.eu