.-∅

D2.4.: Beyond risk and harm – issues of XR technologies and the need for ethics-oriented design and development

Authors: Cristiana-Anca Voinov, Rosemarie de la Cruz Bernabe

Editor:

Project title: The Equitable, Inclusive, and Human-Centered XR Project

Project acronym: XR4Human

Funded by the European Union

Deliverable factsheet:

Project Number:	101070155
Project Acronym:	XR4Human
Project Title:	The Equitable, Inclusive, and Human-Centered XR Project
Title of Deliverable:	D2.4. Beyond risk and harm – issues of XR technologies and the need for ethics-oriented design and development
Work Package:	WP 2
Due date according to contract:	M27
Editor(s):	
Authors(s):	Cristiana-Anca Voinov (Part 1), Rosemarie de la Cruz Bernabe (Part 2)
Reviewer(s):	Shereen Cox, Bjørn Morten Hofmann, Rigmor Baraas
Approved by	Rigmor Baraas

ABSTRACT:	
Keyword List:	

Funded by the European Union

Consortium:

	ROLE	NAME	Short Name	Country
1.	Coordinator	UNIVERSITETET I SOROST-NORGE	USN	NO
2.	Partner	XR4EUROPE	XR4Europe	BE
3.	Partner	UNIVERSITETET I OSLO	UiO	NO
4.	Partner	ETHNICON METSOVION POLYTECHNION	NTUA	EL
5.	Partner	NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET NTNU	NTNU	NO
6.	Partner	UNIVERSITEIT LEIDEN	ULEI	NL
7.	Partner	KARLSRUHER INSTITUT FUER TECHNOLOGIE	KIT	DE
8.	Partner	INSTITUTT FOR ENERGITEKNIKK	IFE	NO
9.	Partner	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	Fraunhofe	DE
10.	Partner	OPEN AR CLOUD EUROPE GEMEINNUTZIGE UG	OARC EU	DE
11.	Partner	STICHTING FONTYS	Fontys	NL
12.	AP	UNIVERSITY OF LEEDS	UNIVLEEDS	UK

Revision history:

VERSION	DATE	Revised by	Reason
0.1			
1.0			

Table of contents

Introduction 5
Defining XR and other terms 6
Overview and Clarifications 7
PART 1: Problems of Ontology, Epistemology and Ethics 7
Problem 1: Are virtual worlds real? 7
Virtual worlds are not real 8
Virtual worlds are real 9
Some objects/events are real in virtual worlds, others are not 9
Problem 2: Who am I in virtual worlds? 9
Virtual worlds produce multiple selves 10
Virtual worlds create uncertain self-epistemologies 11
Problem 3: How meaningful are virtual worlds? 11
Problem 4: How do I make judgments in virtual worlds? 13
Categories are muddier in VWs 13
Are moral judgements the same in VWs? 14
References 17
PART 2: Ethically Sensitive XR Technology Development: Towards a Normative Framework 21
The Moral Responsibility of Developers to Develop Ethically Sensitive XRTs 21
Ethically sensitive XRT 22
Ethics by Design 22
Implications for XRT development 24
Value-Sensitive Design 24
Implications for XRT Development 25
Human Development Approach 26
Implications for XRT Development 27
Conclusion: The Responsibility of Developers in Creating Ethically Sensitive XRTs 28
References 29
APPENDIX 30

Beyond risk and harm – issues of XR technologies and the need for ethics-oriented design and development

Introduction

EXtended Reality technologies (XRTs) have been referred to as the "future of how humans will interact with technology". (World Economic Forum 2025)[1] These technologies represent a new frontier which marks a shift from 2D interaction to 3D, enhanced visualisation capabilities, and collaborative innovation. Uses for XRTs span across various sectors including but not limited to education, aerospace, manufacturing, and entertainment. XR's affordances (and affordability) is being designed to meet the needs of both the layperson and the expert. Even if XR does not completely transform human-technology interactions, it is evident that it offers unique and useful solutions.

It is critical, however, to establish the ways in which the technology impacts individuals and societies. A recent publication by Voinov et al. (2024) offer a typology of risks based on mapping over 200 risks and harms associated with the design, implementation and use of XRTs.[2] This paper expands upon this typology by addressing issues associated with XRTs which *cannot* be framed or conceptualised in terms of risks or harms. It problematises these issues as questions relating to the ontological and epistemological status of XR environments. In other words, there are some features of XR which, as philosopher Philip Brey points out, "transcend particular uses of them". (1999, 5) These issues, we purport, arise from features inherent to XRTs such as *computer generated-ness, interactivity, and immersiveness*. (Heim 1998; Brey 2008; Chalmers 2017)

Some XR experiences remove the space between viewer and technology – unavoidable in traditional visual media like television or cinema screens. A user is then able to *interact* with objects, entities and other users within a computer-generated environment, having a direct or indirect effect on the environment itself, while also giving the user a sense that they are *immersed* within that space.[3] There is reasonable ground to assume XRTs will be increasingly integrated into everyday life. A complete understanding of XR problematics, as related to these inherent and unique features, is therefore important not only as an exercise in knowledge building but also practically: it has design and practice implications.

This paper is written as part of a series of documents/reports in the XR4Human project within Work Package 2 (WP2). The WP's task is to map, examine and analyse the ethical and philosophical dimensions of XR (D.2.2 and 2.4), as well as any guidelines and normative frameworks pertaining to the governance and management of these issues (D2.1, D2.3, D2.4). These understandings will then

[1] Terms such as XR and its related constituents will be defined in the next section.
[2] The themes and subthemes iterated in this typology include: 1) health and wellbeing (cognitive, psycho-emotional, socio-relational and physical), 2) autonomy (privacy, self-sovereignty, surveillance, authenticity), 3) epistemic and scientific (epistemic uncertainty, scientific uncertainty, algorithmic imperfection, manipulation), 4) normative (norm violations, rights violations), 5) structural (socio-political, economic/financial, legal), and 6) technical safety and security (security, dual use). (Voinov et al. 2024, 28)
[3] Note that we will use physical reality/life and non-digital reality/life as synonymous terms.

Funded by the European Union

underpin a code of conduct for XR developers, in addition to a rating system for XR end-users. A strong knowledge foundation of the fundamental ethical/philosophical issues related to the design, development and management of the technologies is critical, not only for robust policy making but also to facilitate value-centred strategies for those intimately involved in the design and operationalisation of these novel technologies. Finally, it is important to critically reflect on the mediating role these technologies may play within individual lives as meaning-making and identity-shaping entities.

Defining XR and other terms

"XR" is not a technology in and of itself, but rather a term which refers to a wide variety of experiences mediated through various hardware and software. (XRSI 2025) These experiences are united insofar as they blur the line between digital spaces and non-digital spaces, facilitating altered or alternative "realities" which a person can engage with, be it "virtual" (VR), "augmented" (AR), or something in between (mixed; MR).[4] XR, thereby, is a "continuum", wherein non-digital reality is found on one end, and complete virtuality within which space and objects are completely digitised – environments are completely virtual (VR) – are found on the other end. (Milgram and Kishino 1994) This continuum may also be understood in terms of level of *computer-generation*, *interactivity* and *immersion*. (Chalmers 2017; Brey 2008; Heim 1998)

XR technologies offer *computer-generated* environments, which are then transmitted to the user via various channels like screens or Head Mounted Displays (HMDs). All generated environments within the continuum are also *interactive*, in that the user can affect their virtual environment – a user in VR, AR and MR can make a difference to the environments they inhabit. Complicating definitions of XR, however, are rather heterogenous understandings of *immersion*. (Lee 2020) The concept is sometimes defined in subjective terms, "presence", where a person feels like they are "there" in virtual space from a first-person's perspective. More contemporary definitions refer to immersion as an artifact's objective capacity to simulate environments which induce a first-person experience (through advanced sensory range, visual fidelity, etc.). By this latter, objective definition, HMDs offer immersive capabilities, whilst desktops and mobile platforms do not. However, considering that some users do not experience immersion even within a technically immersive environment, the more subjective notions of feeling present are still important to consider. Sometimes these issues relate to the technology not accommodating unique user needs (differences in vision, for example). (Pladere et al 2022) Other times, however, the lack of immersion is related to phenomenological or psycho-socio criteria which are difficult to assess empirically – i.e. a user never really reports feeling "there", even when objective conditions are in place. (Nilsson et al 2016; Lee 2020) Immersion, therefore, is intimately related to both hardware and software capabilities - some platforms are more technically immersive than others (HMDs versus desktop and mobile platforms) – while nevertheless being bound to user experience.

Given the diversity of readers which may access this paper, many of whom are unfamiliar with terms in philosophy, it is prudent to spend some time defining the following: "ontology", "epistemology", and "phenomenology". An ontology is a system of answers to the question of "what is", and in doing so distinguishes from "what is not". Consider a painting of Keanu Reeves versus a photograph of him.

[4] Augmented reality places digital information onto non-digital structures; the "real" world is still central, but digital information supplements it (Pokémon GO, for example). Mixed reality technology allows for the manipulation of both digital and non-digital elements (you may use a non-digital item to move a digital one, for example). (XRSI 2025)

Funded by the European Union

Which Reeves is real? Perhaps both entities are real, but differently real. The Reeves who we met on a chance encounter in New York City, the memories gained from this meeting – these versions of Reeves might be "real", in different ways. An ontology accounts for the ways in which all these Reeves are – or are not, perhaps – real. Epistemology concerns itself with systems of knowledge, truth, and belief. How does one tell the difference between a statue of Keanu Reeves and the person himself? Phenomenology concerns itself with describing subjective experience: what does meeting Reeves feel like? Does recalling meeting him feel the same as the meeting itself?[5]

Overview and Clarifications

This paper both describes what is in the philosophical literature, referring largely to seminal texts in the field of philosophy of technology, specifically those relating to the critical study of VR, AR, MR, virtuality, and human-computer relations. Relevant literature sourced in D2.2's extensive review was also used. (see Voinov et al 2024) It also describes seminal normative frameworks applicable to the (normative) issues iterated. In this way, this text serves as a philosophical overview and critical normative analysis, presenting an argument for the need for ethically sensitive XR technologies. It isorganised into two main parts: 1) Problems of ontology, epistemology and ethics and 2) ethically sensitive technology development: towards a normative framework. Part 1 will detail potential philosophical and normative issues with XRTs which cannot be framed in terms of risks and harms. The section aims to describe the problems as such and will not offer proposed solutions to the problems at hand. The second part will iterate strategies for mitigating some of these issues, as well as issues identified in D2.1 and D2.2., with particular attention to normative problems. It will examine these problems through the lens of three normative frameworks: Ethics by Design, Value Sensitive Design and the Human Development Approach (Van den Hoven 2012; Oosterlaken 2014; Davis and Nathan 2015).[6]

This article will generally focus on the unique offerings of *immersive* experiences, where changes/movements from a user create change within the virtual or augmented environment or *virtual world* from a first-person perspective. As such, whenever the phrase "virtual world" (VW) is discussed hereafter, it will mean any XR experience which is computer generated, interactive *and* immersive. To reiterate, immersion includes both technical components and subjective experience. Given the plethora of AR and MR experiences which can also be immersive, a virtual world may be facilitated via VR, AR or MR.[7] We will use "non-immersive virtual worlds" referring to interactive, computer-generated environments within the XRT spectrum which do not meet the criteria of immersion. In this way, we acknowledge that non-immersive XR may also engage feelings of presence, and thus may warrant similar (though not identical) philosophical conclusions, while still maintaining a conceptual focus on immersive features of XRTs.

PART 1: Problems of Ontology, Epistemology and Ethics

Problem 1: Are virtual worlds real?

Fully immersive VR facilitated through XR technologies is sometimes referred to as aiming towards an "essential copy" of non-digital life. (Bown et al 2017) This "essential copy", in its idealised form,

[5] For in-depth discussions see Hofweber (2023) (ontology), Steup and Neta (2024) (epistemology), and Smith (2013) (phenomenology).
[6] There are multiple frameworks which would be applicable here. For practical purposes three were selected.
[7] As such, the focus will not be on XR devices but rather their facilitated experiences. We will also clarify when a problem relates singularly to a VR, MR or AR experience.

Funded by the European Union

would become an "Ultimate Display", rendering a fully digitalised, indistinguishable replica of non-digital space. Though this Matrix-like[8] scenario does not exist, as Brey correctly reflects, there is a "virtual or electronic equivalent of almost anything found in the physical world" within virtual environments (both immersive and non). (Brey 2003, 269) Indeed, there are virtual equivalents for most objects, but also events – virtual avatars may host a wedding, play tennis, or share a meal, for example.

It seems, however, that these objects/events are not strictly "equivalent" to those of the physical world – these objects and events are made of digital bits, while non-digital objects are made of atoms. The possibilities and physical properties afforded by digital things to those of the physical world are not, at least materially, the same. They are *computer generated*. Additionally, the events or objects found in virtual environments do not have to correspond to that of the non-digital – they can be made up of completely imagined entities (fantastical lands, alien flora and fauna, magic, etc.). They also can be files, virtual documents, and so on.

With the expanding range of XR devices, users interact with virtual environments and perceive their objects. This is intensified in VR where users experience embodiment and presence within virtual space – they feel like *they* are *there,* in a *virtual* body. "In effect, being computer-generated makes these environments virtual… and being immersive and interactive makes our experience of them *at least akin* to ordinary reality". (Chalmers 2017, 3; emphasis ours) Does the fact of computer generation make a difference to whether an object or interaction in virtual worlds are real? What is the fundamental nature of these virtual worlds?[9]

Virtual worlds are not real

Some scholars purport that VWs, and therefore objects and interactions within them, are *not* real. This view is often predicated on realness being mind-independent, or "independent of the way we think, talk, experience, or conceptualise". (McDonnell and Wildman 2019, 372) By this understanding, the entities which a VW inhabitant experiences are fictional – they do not exist outside of their conceptualisation and experience.

Virtual objects are fictional objects. Ordinary objects have a set of specific structures which provoke their status as that object. Water, for example, is water because it is made up of hydrogen and oxygen – that is how we may differentiate water from, for example, methanol, which at first glance behaves and appears to be water-like. This does not hold for virtual objects. Additionally, virtual events are argued to be perceptual illusions of some kind. Here, VW's trigger a perceptive response which make a user believe in the existence of an occurrence as a non-virtual occurrence, but that response is based on a hallucination.

This is, notably, how technologists like Slater (2009) talk about immersive experiences, referring to the "plausibility *illusion*" and "place *illusion*", where events taking place in virtual worlds and especially VR are convincing portrayals but ultimately akin to a mind-trick of some sort. "As part of

[8] A «Matrix-like scenario» refers to the 1999 science fiction film, *The Matrix*, in which the protagonist uncovers that he has been living in a perfect simulation, where "reality" itself is computer generated. We will not be exploring these rather fantastical understandings of XR, and will remain focused on how XR is currently used, or will probably be used in the near future.

[9] The positions offered here are simplified forms of each argument. There are different types of realist and irrealist positions, predicated on different assumptions about, for example, perceptual independents of objects, etc. (McDonnell and Wildman 2019)

Funded by the European Union

the Ultimate Display" write Bown et al (2017), "an essential copy can be thought of as a perfect rendition of an object that *fools* all sense into perceiving it as real" (emphasis ours).

Virtual worlds are real

Conversely, others contend that virtual worlds are, indeed, real. These are realists such as Michael Heim (1998) and David Chalmers (2017), who purport (though not always simultaneously) that virtual objects really exist, that interactions are really occurring, and that while illusions can take place in them, the world's fundamental structure is non-illusory.

One way to support the realist position is to say that the underlying structure of these worlds are real, just made up of different material: computational material. This status extends to objects in AR and MR; therefore, it is not exclusive to a fully digitalised world. Interactions take place between distinct bodies (avatars) by way of computational structures within a digitised space. The user experiences, by way of perception, the properties of the structures themselves.

There are, then, causal effects upon the user by way of these structures and their perceivable properties. These underlying structures create a realist fabric to virtual worlds *even if* the scenarios taking place may be fictitious. "If a human in physical reality plays the role of Gandalf casting a spell in Middle Earth, the event of Gandalf casting a spell is fictional, but the underlying bodies and movements are real". (Ibid., 6)

Realists like Chalmers argue that virtual properties can be equated with their experiential effect, or how they function to the perceiver. Leaves in physical reality are green because they are experienced as green.[10] Similarly, "the data structure corresponding to a virtual red rose really does cause reddish experiences… so the data structure is virtual red… even though it is not non-virtually red". (2017, 12) Virtual objects and events are not the same as non-virtual objects or events (though they can be), but both are real. The user therefore experiences a real virtual object, or interaction – even if that virtual object is positioned in physical space as in the case of MR.

Some objects/events are real in virtual worlds, others are not

There is also a middle ground position, offered by philosophers like Philip Brey and John Searle, who qualify some objects, events or experiences as real (Brey 2003, 2008; Searle 1995). Brey separates virtual entities into "simulations" and "ontological reproductions", wherein the former is perceptually like their non-digital counterpart but does not contain the material or practical components which would allow it to behave like the real-world equivalent. (2008) A virtual apple can look like an apple, and with convincing haptic technology may even feel like an apple, but you cannot sustain yourself with it, for it is not made up of the stuff that makes up an apple. On the other hand, there are ontological reproductions which, by way of collective agreement or institutional practice, are made real. A user is real, and whilst their eating a simulated apple is not real, their interactions with other users as embodied and present agents, makes certain acts real. This is considered a sort of "pragmatic" distinction. (Ibid.)

Problem 2: Who am I in virtual worlds?[11]

It has already been stated that some XRTs, unlike previous media, place the user in a position of action, wherein they are able to interact with, and be immersed within, a virtual world. Is the user,

[10] For a colour normal person.
[11] Additionally, who am I *with* virtual worlds?

Funded by the European Union

however, a representational entity? Are the virtual and physical distinct selves? In addition to the questions brought about by the immersive and interactive affordances of these technologies, non-immersive XR also may bring about changes to individuals' identities outside of the virtual worlds themselves, for example, by persistent wearability and bodily integration. This also brings forth new ways of understanding personal identity: are humans singularly biological, or more than that? This section discusses the following issues: the problem of more-than-one selves, and the problem of a potentially distorted epistemology of the self.

Virtual worlds produce multiple selves

VR is often talked about as a technology which facilitates more-than-one "selves". "Technological VR is the representation of *possible* worlds and *possible* selves, with the aim of making them appear ever more realistic". (Metzinger 2008, 3) These possible selves seem predicated by being able to identify "as a self" within a specific time (x) and space. (Ibid.) What possible "selves" refers to is rather puzzling. An avatar is not an exact replica of an individual. It is not physically continuous to the self in physical reality; an avatar does not possess the same brain.

Perhaps what VR gives rise to rather akin to a second *life* – the IEEE's initiative reporting on the Ethics of Extended Reality (XR), for example, refer explicitly to "the ownership of second lives" within immersive tech. (Cortese and Outlaw 2021, 5) This is not meant literally, but rather refers to the creation of a parallel but supplemental existence which may hold equivalent psycho-socio meaning to that of physical reality, without the material conditions of physical reality which are necessary to life itself (food, water, etc.). An individual's personality, their preferences, their social relationships, are transferred across material mediums.

While this might be the case in non-immersive virtual environments, it would seem that VR *does* create a second self by way of "deindividuation" which then leads to a "Proteus Effect" phenomenon. "Deindividuation", defined by McKenna and Bargh (2000, 61) is "when an individual's self-awareness is blocked or seriously reduced by environmental conditions" resulting in "a weakened ability for an individual to regulate his or her own behaviour, reduced ability to engage in rational, long-term planning, and a tendency to react to immediate cues or based largely on his or her current emotional state. Furthermore, an individual will be less likely to care what others think of his or her behaviour and may even have a reduced awareness of what others have said or done". On this view VR can at least inhibit selfhood. The notion of a second self comes in where the avatar begins to *replace*, to some extent, a deindividuated user's identity. (Yee et al. 2009) It is not clear, however, if the experience of being an avatar alters features of myself or the "I" in its entirety, referred to by Limanowski (2014, 2) as the "subject of experience".

In addition to creating new in-world subjectivities for the user, being in virtual worlds may restructure selfhood in the physical world. It may not only be the case that Proteus effects extend into real life, and therefore *presentations* of selfhood in physical reality become altered (as with VR) but it may also be the case that VR, MR and AR may change the way the self is understood and related to in physical reality. Featherstone (in Cui 2016, 170) sees this as a function of blurring analytical categories: "The key analytical categories we have long use to structure our world… categories of the biological, the technological, the natural, the artificial and the human are now beginning to blur". Perhaps this can be better referred to as the restructuring of a once purely flesh-person into a cyborg – part-human, part machine. It is worth noting that philosopher and biologist Donna Haraway (2016, 7), among others, have argued that the "hybrid of machine and organism… a fusion of the organic and the technical" have been long forged in "particular, historical, cultural practices". In other words, boundaries have always been blurred – humans and technologies have mediated and co-constituted the self. Still, as

Funded by the European Union

wearables become increasingly easier to use, more seamlessly integrated into everyday life, and persistently worn (no need to take them off), our understanding of what it means to be a physical self will change in the near future.

Finally, the *phenomenal* model of the physical self is not transferred to the avatar upon entering a VW. (Metzinger 2018; Madary and Metzinger 2016) That is, selfhood is indeed both an "objective" physical thing but also a partly subjective experiential thing. (Limanowski 2014) Between the physical world and its immersive virtual counterpart, there is a phenomenal break between physical self and virtual self, or perhaps the creation of distinct phenomenal agents: the physical entity/body and the virtual entity/body. (Chalmers 2017)

Virtual worlds create uncertain self-epistemologies

Another problem that may arise from the use of VWs is the possibility that they, may create uncertainty in terms of whether a person can be "fully themselves" within immersive spaces. This question is partially answered in Problem 1 as a function of phenomenology and ontology, but also is about *determinism* versus *free will*. Virtual worlds are artificial spaces beholden (though not entirely) to the system designer(s)' input. They are therefore externally controlled or determined. This problem is particularly salient in VR, where worlds are completely digitised. In what sense is a user's avatar in control of their own choices?[12] It is argued that the surveilled, artificial construction of these worlds means the mental integrity of a person – their right to control their mental state and the contents thereof – is vulnerable (at least in comparison to physical reality). (Ligthart et al 2021) Maslen and Salvulescu (2018, 590) refer to the related problem of "authenticity". Their focus is not on whether an individual becomes inauthentic, but rather if an individual can express themselves authentically in virtual environments in the first place – whether there is something inherent to the medium itself which does not allow for the full range of expressive selfhood, perhaps because of its artificiality, or something else.

The controlled artificiality, some argue, may additionally lead to various "epistemic injustices" – part problem of knowledge, part ethical issue. (Scotto 2020) This kind of injustice is not about lacking access to knowledge (a problem of distribution) but is rather about how systems of power may alter or skew the way another person ascertains knowledge (not always intentionally) in a way that disadvantages or harms that person. (Scotto 2020) Epistemic injustices may also alter the way people come to know, or understand, entire groups of persons. (Ibid.) How may VWs – often constructed and embedded with the knowledge structures of white, able-bodied men – cultivate versions of you which may limit other user's knowledge of you, but also "of their own social experience" – especially when such environments are fully immersive? (Scotto 2020, 157)

Problem 3: How meaningful are virtual worlds?

Questions of meaning are critical to consider, as user bases for mobile AR alone reach over 1 billion persons across the world, and time spent using XRTs increases (Statista 2024). There are several "threats" to meaning which scholars identify, including the threat from: lacking quality, being artificial, having no history (argument from transience), being intrusive/a distraction, facilitating separation, lacking embodiment, and facilitating indefinite life.

1) *VWs lack quality*. Simulated environments, including non-immersive, are expressed in lower fidelity to the physical world; simulations are imperfect (or bad). The precision or exactitude

[12] With the introduction of AI assisted "hybrid" avatars, wherein partial control is afforded to the user and supplemental control to the AI, the question becomes all the more complex.

Funded by the European Union

in which environments can be recreated, or in which faces and movements can be expressed are lower. This may have an impact on the way emotion is conveyed, how social cues are gartered, or simply on how persons experience their simulated surroundings.

2) *VWs are artificial.* Some liken VR, in particular, to Robert Nozick's (1974) "Experience Machine". In his thought experiment, a machine offers the capacity to give a user any desired experience by way of "plugging in", leaving their body on standby while their brain wholly believes itself to be within that experience. Nozick asserts that despite this desire fulfilment, a user should not choose to plug in. He reasons that people want to be active doers, facilitators of their own lives, and not just passive experiencers. Physical reality offers choice and chance – maybe a user does not get what they want – but in exchange it offers the capacity for personal involvement, leading to a richer life. A machine of artificial human construction and control does not offer the depth of the physical reality. In Nozick's view, this kind of machine is illusory and pre-determined.[13]
3) *VWs have no history.* On this view, physical reality is meaningful because it has had a long history and will (hopefully) go on to have a lengthy future. Virtual worlds, including non-immersive, are brief, they have not had lengthy pasts and arguably will not have lengthy futures. Arguably, the virtual world itself only exists if there are one or more user/s in it. The physical world, in contrast, continues to exist – undergoing wear and tear with or without humans.
4) *VWs are intrusive.* This argument relates to how VWs, both immersive and non-immersive, impact or distract a person from their physical life. There are several ways in which this may happen (see Voinov et al 2024), but in general this can mean that XRTs may have an impact on a user's normal behaviour by way of distracting them, creating addictive patterns, or leading them to neglect themselves or their non-digital social/familiar worlds. Additionally, some posit not only an intrusion, but that VR has been reported to facilitate a behaviour transfer from simulation to reality. (Slater et al. 2020)
5) *VWs facilitate separation.* This problem, related closely to the previous, is that VWs offer two separate worlds for someone to live in. The more immersive, the more embedded and reality-like the virtual world becomes. This duality explicitly assumes the primacy of the physical world, wherein the virtual world lacks the social and familial relationships of the former, making it a less valuable reality.
6) *VWs lack embodiment.* There are no *physical* bodies in simulated environments, and a user certainly does not import their own body – even if they experience embodiment within an avatar form.[14] A physical body offers a range of possibilities, not only in terms of motion, but also tactile sensation, smell and taste. Their lack is therefore a value loss. Theorists have a difficult time imagining a Matrix-like scenario where all the body's functionalities can be replicated exactly – especially the body's functions as they are experienced in your own body, with its unique shape, size, capacities.

[13] It is worth noting that comparisons between VR and Nozick's experience machine have been highly critiqued (see Cogburn and Silcox 2014).
[14] This latter point, where one becomes disembodied from their *own* body, will be discussed further below.

Funded by the European Union

7) *VWs lack a virtual analogue to birth and death.* Birth and death, two events which scholars (and ordinary people) arguably attach great value to, do not occur in simulations.[15] "Birth is just registering an "account" and "death" is just a form of "cancellation". (Cui 2016, 181) A virtual life can be indefinite, or at least reoccurring, in XR environments because an avatar can be reconstructed.

These scholars purport that for these reasons, and others, 1) VWs are unable to offer equivalent meaning for users to that of physical reality (arguments 2, 4, 7, 8) or even 2) it may actively negatively impact a user (arguments 5, 6, 1). What some of these arguments hinge on is an evaluation of virtual worlds in their current technological capacity. However, technology may improve fidelity and offer more ways to incorporate full-body sensory experiences. Other arguments centre meaning around events or features which cannot be replicated: physical reality's history, its naturalness, and the fact that persons can be born and die. At best, some say, virtual worlds can offer a role in meaning-making akin to that offered by a film or book.

Problem 4: How do I make judgments in virtual worlds?

A main concern within philosophy is that of knowledge formation (Ichikawa 2017). To *judge* something is to make an assertion about the truth of that thing; it is, in part, a knowledge assertion (in that it requires knowledge) (Ibid.). If we take knowledge to be a *justified true belief* about something, we find that the ontological muddiness of virtual worlds makes it more difficult to assess whether something is justified, or indeed what truth means. (Metzinger 2018) *Justification*, according to scholars, entails evidentiary support, perhaps in the form of observation (sensory input) or testimony (appeals to expertise). *Truth* may be ascertained by correspondence to reality, or deduced logically from a set of axioms, or by consensus, or perhaps by what is useful. (Ichikawa 2017) Yet, wholly relying on our senses to determine what is true or false in virtual worlds is problematic, and so is determining whose expertise to rely on (an avatar? Another user? A designer?).[16] What does "correspondence" even mean in a virtual world, and what kind of pragmatic consensus-building can take place in one?

There are several issues to cover here: the problem of category muddiness which highlights the issues in how self-knowledge and self-image is formed, how to distinguish virtual objects from physical objects in AR and MR, as well as problems derived from mutable "boundaries" in virtual worlds. These can be seen as issues with knowing *inside* of virtual worlds, including knowing or determining how to distinguish *between* digital and physical. Finally, this section will discuss ethical judgement-making in virtual worlds.

Categories are muddier in VWs

This blurriness can be understood in two ways. First, the mutable structure of simulated environments requires a learning of "natural laws" within each virtual world – who and what can I do, but also what other entities can do (internal muddiness). For example, the category of "limits" is integral to how persons within the physical world create a set of expectations for the self, but also between other persons, objects, events, etc. In the physical world, a person is certain (more or less) of their physical limits; they know their own bodies cannot take flight, they know they cannot outrun a car on the

[15] Unless sentient artificial minds are posited, which may theoretically be born and die within virtual environments.

[16] This is not to say that we can entirely rely on our senses in physical reality. But sense-perception is one strong evidentiary tool in physical reality. Sensory inputs in VR, however, are more complex. What other means towards knowing do we have in VR?

Funded by the European Union

highway. A user of virtual worlds understands some limits insofar as software and hardware capacities, and typical in-world socialised behaviours, but expectations between users are not "fixed" in the way physical reality is. Therefore, virtual worlds require a learning of "limits", and sometimes a "re-learning" of the laws (material and normative) which bracket how we interact both physically and socially with persons. "Who gets to decide what these [designs] look like, and how easily can they change"? (Voinov 2024, 137)

A user's epistemic foundation is arguably shakier compared to that of non-digital reality. The reliance on, for example, expert testimony or even a user's senses is less secure. Indeed, a face-to-face meeting in non-digital reality is usually enough to warrant forming a justified true belief about their existence – or that they are a person and not a complex AI, or that they are the person they say they are and not someone else. This is not the case in VWs:[17]

> "What would be a non-hackable mechanism for reliably identifying the current user(s) of a given avatar? But even if we find such a mechanism, the epistemological problem of other minds remain. Even if I can be convinced of the identity of an agent I encounter in VR in a way that suffices for all practical and legal purposes, I will still be interested in a high degree of certainty when it comes to more direct interpersonal relationships in social VR" (Metzinger 2018, 8)

Even if VWs are ontologically real, it can still be the case that illusions, and convincing ones, will take place in VWs. External tampering or hacking makes this a possibility. (Chalmers 2017) These tampering efforts are much easier to induce than illusions in physical reality.

A final point to consider is that there is perceptual muddiness in distinguishing between the category or digital and physical itself (inter-reality muddiness). While perceiving objects and events within VR seems more straightforward, is it the same for virtual objects in AR or MR, where an intermingling of realities occurs? Here, advises Chalmers, it becomes helpful to shift the analysis away from perception and onto belief. (2017) Users generally do not *believe* that objects in VR are nonvirtual, rather, they are aware that they are interacting with a virtual object or being.[18] (Chalmers 2017) Similarly, virtual objects/entities in MR and AR are identifiable as virtual objects, and therefore will be interacted with as such. These interactions may be interpreted as taking place in a single space or plane (physical reality, for example), but the object themselves will be identifiable as such. Problems arise when (and if) it becomes difficult to identify a virtual object as such because it is designed to behave like a non-digital object/entity. Additional problems arise when witnessing objects which are part virtual, and part physical, where an object being used in physical space is represented in virtual space.

Are moral judgements the same in VWs?[19]

One understanding of moral judgement is that it is an assertion made pertaining to the rightness or wrongness of an act. (Hooker 2017) A moral judgement can be seen as a form of reasoning, involving factors including ethical/moral theory, principles, affect/emotion, social and cultural context, among other things. Harassment or violence against a person or animal is generally considered a prima facie moral wrong in the real world, for example, but are these moral wrongs also applicable to harassment

[17] To be clear: this is not a denial that justified true beliefs are not possible in VR, just that they are perhaps more difficult to form, or will require new (different? Additional?) modes of justification.
[18] Technologists have not yet achieved the Ultimate Display as such.
[19] Ethical judgement and moral judgement are used synonymously. Additionally, in this paper moral judgements are a subset of normative judgement, rather than synonymous, as normative judgements also include reasoning based on rules of etiquette or law, for example.

Funded by the European Union

or violence in virtual worlds (immersive and non-immersive alike)? Given an increasingly weighty body of evidence suggesting that avatar harassment and assault (particularly in VR) can be experienced similarly to violence in physical reality (Slater et al 2020), it would seem that virtual violence should be judged similarly to physical violence. Yet, even if virtual worlds are "akin" to reality, to echo Chalmers (2017), there does seem to be morally-relevant differences in who is judged as a moral actor, what tools are used to judge them and the appropriateness of the tool, and what type of consequences are ascribed to the violator(s).

That is not to say that people will always choose to forgo moral reasoning or moral action in virtual spaces. There is no denying that there are social constraints and rules which manage in-world behaviour, or that there are consequences for acting wrongly. (see Huff et al 2003) Yet, it is nevertheless difficult to assess wrongness on equivalent terms across physical-to-virtual mediums. The seminal 1993 "rape in cyberspace" case in multi-user virtual environment LambdaMOO illustrates this. (MacKinnon 1997) The case features a character called Bungle, who takes control of two avatars and forces them to perform sexually violent and otherwise deviant acts on other players. Authors Huff et al. reflect that intuition seems to indicate Bungle's assault to another avatar is not strictly the same thing as an assault in physical reality. However, they go on to purport that there does seem to be something "wrong" about Bungle's actions. The authors go on to write that "our ideas about sex and harm are deeply tied to flesh. The cyber-space rape is puzzling because it suggests that rape can happen to characters that are not flesh." (Huff et al 2003, 14) Indeed, their suggestion implies that either an avatar body is itself a body which can be harmed, or perhaps that the moral wrongness of such cases could be ascertained on grounds not tied to flesh, but are rather about coercion, exploitation and control as violations of autonomy or dignity (among other things).

Yet, direct consequences to the flesh do play a significant role in moral judgement-making in physical reality. To illustrate this, we can refer to the classic example of "the trolley problem". (see Foot 1967)[20] According to Foot, it would seem permissible to kill one track worker by deliberately redirecting a troller away from hitting five workers. It is hard to image that the trolley problem would hold the same moral weight in virtual reality. Even if the people on the track experienced the fear or emotional harm in believing that their death is imminent, there is no real death which takes place in VWs. The *stakes* are different. Nor are avatars punished in the same way punishment is enacted in physical reality. The worst that can happen is the perpetrator gets suspended from the world itself.

Another way of determining how to make a moral judgement would be to start with the question: Is a person's avatar and a person the same? A person is considered a moral agent, worthy of moral consideration. If a person and their avatar is equivalent, a person's avatar should be treated in the same way as their physical selves. We may then, for example, apply axioms like Kant's "categorical imperative", wherein we treat the avatar like an end-in-itself. (Kant 1998 [1785]) However, Adomaitis et al. 2020 argue: "Material people… cannot *meaningfully* put themselves in the avatar's place. Therefore, ethical behaviour in a metaverse deserves a different ethical treatment not based on projecting the categorical imperative on avatars, despite the fact that there are real psychological effect of the behaviour happening in the metaverse". (2020, 42, emphasis ours)[21] The authors do not say what they mean by "meaningfully"; perhaps they mean that the way something is perceived by a

[20] The trolley problem is a thought experiment designed to probe the permissibility of actions which involve knowingly harming person(s) as a byproduct of a deliberate action.
[21] The moral concerns of a synthetic character, an AI controlled avatar for example, would also warrant different treatment. Perhaps users could treat them as a mere means.

Funded by the European Union

person (phenomenology), and the metaphysics of the world itself (ontology), are not *the same* between realities (even if we assume both experiences to be ontologically real).

What about the moral status of an avatar's stuff – or perhaps the virtual objects belonging to a real user? Theft or vandalism in VWs seems to be worthy of moral concern, but not the same as similar violations in physical reality. Perhaps their moral status is, again, not tied to an object's materialism, but rather as a function of violations to user autonomy or dignity.

One conclusion would be to reformulate strategies for ethical decision making based on the unique phenomenology and material conditions of XR – the establishment of XR specific moral (and normative) theories which arise out of the milieu(s) themselves.[22] This is a difficult task, as it is arguably impossible to totally divorce a user's moral intuitions or reasoning from their understandings of the physical reality in which they normally "live". Here, we return to the concept of a "moral agent". Not only are moral agents worthy of moral consideration, but they are also those who can be "expected to meet the demands of morality" (Routledge Encyclopaedia of Philosophy 2024). Given that the demands of morality may reasonably look different in VWs, so might identifying who is an agent and who is not. Could an agent be an AI-avatar? A strong interpretation of Kantian ethics, for example, would argue yes. If an agent can deduce the moral law and act in accordance with it, it is a moral agent. Emotions (according to a Kantian) need not necessitate a moral agent – an AI avatar might operate under a deductive logic and execute actions accordingly.

What about the role of designers and developers as agents? Unlike physical reality, VWs are designed by developers who have a say in the possibilities of each world – they can control what an avatar does and does not do. Thereby, would it not be the case that a programmer(s) can be at least partially responsible for moral wrongs which occur in spaces designed by them? This is further complicated because complex programming can create outcomes in virtual spaces which the programmer 1) did not intend and 2) could not always foresee.

If, as illustrated thus far, moral theory, moral principles, or moral agency is difficult to determine in VWs, a similar difficulty would be determining *moral virtues*. Virtue ethics purports that cultivating morality is a function of cultivating "good" virtues, by which an agent is thereby able to live a "good" life. (Hursthouse 2022) Some virtues are perhaps easier to identify in simulated spaces (compassion, justice). However, some virtues are more difficult to identify and perhaps even cultivate, such as "courage". What makes a person truly courageous in XR? What is courageous about stepping in to prevent a fight between avatars? Perhaps there are emotional stakes. However, physical stakes in virtual spaces are lower – after all, an avatar can respawn.

A final consideration, though rather indirectly related to judgement-making in VWs, is moral transference from virtual environments to real-world contexts. This refers to not only moral structures and conditions but also norms more generally slipping between virtual and physical spaces. (Brey 1999) Perhaps in some cases the slippage is desirable, but it may be a problem if certain cases. For example, it is posited that VR more easily allows users to engage with violence and harassment – with fewer consequences (certainly physical but also social). (Metzinger 2018) It is unclear how this engagement may lead to moral apathy about those behaviour, or a reestablishment of norms themselves.

[22] Does this call for ethical theorists and philosophers of technology to "live" or "practice" within XR?

Funded by the European Union

References

Adomaitis, Laurynas, Alexei Grinbaum, and Dominic Lenzi. 2020. 'TechEthos D2.2: Identification and Specification of Potential Ethical Issues and Impacts and Analysis of Ethical Issues of Digital Extended Reality, Neurotechnologies, and Climate Engineering'.

Bown, Johnathan, Elisa White, and Akshya Boopalan. 2017. 'Looking for the Ultimate Display: A Brief History of Virtual Reality'. In *Boundaries of Self and Reality Online*, edited by Jayne Gackenbach and Johnathan Bown, 239–59. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-804157-4.00012-8.

Brey, Philip. 1999. 'The Ethics of Representation and Action in Virtual Reality'. *Ethics and Information Technology* 1 (1): 5–14. https://doi.org/10.1023/A:1010069907461.

———. 2003. 'The Social Ontology of Virtual Environments'. *The American Journal of Economics and Sociology* 62 (1): 269–82.

———. 2005. 'Virtual Reality'. 2005. https://www.4tu.nl/ethics/downloads/default/files/brey-2005-virtual-reality.pdf.

———. 2008. 'Virtual Reality and Computer Simulation'. 2008. https://www.4tu.nl/ethics/downloads/default/files/brey-2008-vr-cs.pdf.

Chalmers, David. 2017. 'The Virtual and the Real'. *Disputatio* IX (46). https://doi.org/10.1515/disp-2017-0009.

Cogburn, Jon, and Mark Silcox. 2014. 'Against Brain-in-a-Vatism: On the Value of Virtual Reality'. *Philosophy & Technology* 27 (4): 561–79. https://doi.org/10.1007/s13347-013-0137-4.

Cortese, Michelle, and Jessica Outlaw. 2021. 'The IEEE Global Initiative on Ethics of Extended Reality (XR) Report: Social and Multi-User Spaces in VR: Trolling, Harassment and Online Safety'.

Cui, Shuqin. 2016. 'Cyborg Bodies: Transgression across the Real and the Virtual'. In *Gendered Bodies*, 170–90. Toward a Women's Visual Art in Contemporary China. University of Hawai'i Press. https://www.jstor.org/stable/j.ctvvn7wm.15.

Dubljević, Veljko, Sebastian Sattler, and Eric Racine. 2018. 'Deciphering Moral Intuition: How Agents, Deeds, and Consequences Influence Moral Judgment'. *PLOS ONE* 13 (10): e0204631. https://doi.org/10.1371/journal.pone.0204631.

Haksar, Vinit. 2016. 'Moral Agents'. In *Routledge Encyclopedia of Philosophy*, 1st ed. London: Routledge. https://doi.org/10.4324/9780415249126-L049-1.

Haraway, Donna. 2016. *Manifestly Haraway*. University of Minnesota Press. https://www.sfu.ca/~decaste/OISE/page2/files/HarawayCyborg.pdf.

Hofweber, Thomas. 2023. 'Logic and Ontology'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2023. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/logic-ontology/.

Funded by the European Union

Hursthouse, Rosalind, and Glen Pettigrove. 2023. 'Virtue Ethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2023. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/.

Ichikawa, Jonathan Jenkins, and Matthias Steup. 2024. 'The Analysis of Knowledge'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2024/entries/knowledge-analysis/.

Kant, Immanuel, Mary J. Gregor, and Immanuel Kant. 1998. *Groundwork of the Metaphysics of Morals*. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge Univ. Press.

Lee, Hyunkook. 2020. 'A Conceptual Model of Immersive Experience in Extended Reality' - PsyArXiv preprint. 10.31234/osf.io/sefkh.Limanowski, Jakub. 2014. 'What Can Body Ownership Illusions Tell Us about Minimal Phenomenal Selfhood?' *Frontiers in Human Neuroscience* 8 (November). https://doi.org/10.3389/fnhum.2014.00946.

MacKinnon, Richard. 1997. 'Virtual Rape'. *Journal of Computer-Mediated Communication* 2 (4): JCMC247. https://doi.org/10.1111/j.1083-6101.1997.tb00200.x.

McDonnell, Neil, and Nathan Wildman. 2020. 'Virtual Reality: Digital or Fictional?' *Disputatio* 11 (55): 371–97. https://doi.org/10.2478/disp-2019-0004.

McKenna, Katelyn, and John Bargh. 2000. 'Plan 9 From Cyberspace: The Implications of the Internet for Personality and Social Psychology'. *Personality and Social Psychology Review* 4 (1): 57–75. https://doi.org/10.1207/S15327957PSPR0401_6.

Metzinger, Thomas K. 2018. 'Why Is Virtual Reality Interesting for Philosophers?' *Frontiers in Robotics and AI* 5:101. https://doi.org/10.3389/frobt.2018.00101.

Milgram, Paul, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. 'Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum'. In *Telemanipulator and Telepresence Technologies*, 2351:282–92. SPIE. https://doi.org/10.1117/12.197321.

'Mobile AR Users 2023-2028'. 2025. Statista. Accessed 27 January 2025. https://www.statista.com/statistics/1098630/global-mobile-augmented-reality-ar-users/.

Motilal, Shashi, Keya Maitra, and Prakriti Prajapati. 2021. *The Ethics of Governance: Moral Limits of Policy Decisions*. Singapore: Springer. https://doi.org/10.1007/978-981-16-4043-8.

Nilsson, Niels, Rolf Nordahl, and Stefania Serafin. 2016. 'Immersion Revisited: A Review of Existing Definitions of Immersion and Their Relation to Different Theories of Presence'. *Human Technology* 12 (2): 108–34. https://doi.org/10.17011/ht/urn.201611174652.

Noonan, Harold, and Ben Curtis. 2022. 'Identity'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/identity/.

Origgi, Gloria, and Serena Ciranna. 2017. 'Epistemic Injustice: The Case of Digital Environments'. In *The Routledge Handbook of Epistemic Injustice*. Routledge.

Pladere, Tatjana, Ellen Svarverud, Gunta Krumina, Stuart J. Gilson, and Rigmor C. Baraas. 2022. 'Inclusivity in Stereoscopic XR: Human Vision First'. *Frontiers in Virtual Reality* 3 (November). https://doi.org/10.3389/frvir.2022.1006021.

Robeyns, Ingrid, and Morten Fibieger Byskov. 2023. 'The Capability Approach'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2023. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/capability-approach/.

Royakkers, Lambèr, Jelte Timmer, Linda Kool, and Rinie van Est. 2018. 'Societal and Ethical Issues of Digitization'. *Ethics and Information Technology* 20 (2): 127–42. https://doi.org/10.1007/s10676-018-9452-x.

Scotto, Silvia Carolina. 2020. 'Digital Identities and Epistemic Injustices'. *HUMANA. MENTE Journal of Philosophical Studies* 13 (37): 151–80.

Searle, John R. 1995. *The Construction of Social Reality*. Simon and Schuster.

Slater, Mel. 2009. 'Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1535): 3549–57. https://doi.org/10.1098/rstb.2009.0138.

Slater, Mel, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, et al. 2020. 'The Ethics of Realism in Virtual and Augmented Reality'. *Frontiers in Virtual Reality* 1. https://www.frontiersin.org/articles/10.3389/frvir.2020.00001.

Smith, David Woodruff. 2018. 'Phenomenology'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/phenomenology/.

'Spatial Computing Could Lead the next Big Technology Revolution'. 2025. World Economic Forum. 2025. https://www.weforum.org/videos/spatial-computing-next-technology-revolution/.

Steffensen, Lisa, Kjellrun Hiis Hauge, and Rune Herheim. 2021. 'Ethical Thinking and Programming', September. https://doi.org/10.5281/ZENODO.5416599.

Steup, Matthias, and Ram Neta. 2024. 'Epistemology'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2024. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2024/entries/epistemology/.

Thomasson, Amie. 2022. 'Categories'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/categories/.

Voinov, Cristiana. 2024. 'Bypassing the Body? Immortality and Vulnerability in Digital Spaces'. In *Limits of Life*, edited by Martin Eggen Mogseth and Fartein Hauan Nilsen, 1st ed., 129–40. Reflections on Life, Death, and the Body in the Age of Technoscience. Berghahn Books. https://doi.org/10.2307/jj.9853272.13.

Voinov, Cristiana, Rigmor C Baraas, Rosemarie Bernabe, Shereen Cox, and Lene A. Hagen. 2024. 'XR4Human D2.2: Mapping of Risks and Harms of eXtended Reality (XR) Technologies', October. https://zenodo.org/records/14044988.

'XRSI Standard Taxonomies for Emerging Tech Ecosystem'. n.d. X Reality Safety Intelligence (XRSI). Accessed 27 January 2025. https://xrsi.org/definitions.

Yoh, Myeung-Sook. 2001. 'The Reality of Virtual Reality'. In *Proceedings Seventh International Conference on Virtual Systems and Multimedia*, 666–74. https://doi.org/10.1109/VSMM.2001.969726.

PART 2: Ethically Sensitive XR Technology Development: Towards a Normative Framework

Whether individuals consider XR as adding supplementary meaning to their physical realities, as spaces capable of offering equally meaningful experiences, as ontologically real, or illusory, virtual environments will increasingly play a mediating role in many people's lives. It is worth, then, considering how XR (inclusive of hardware and software) may be designed in a way that is ethically sensitive. Some guidance, therefore, is needed especially in the design and development of XRTs.

The first section detailed issues related to XR which could not be conceptualised as a risk or a harm. The section, in conjunction with previous literature on risks and harms of XRTs developed by the XR4Human team in D2.1., D2.2., and D2.3., is meant to provide the metaphysical as well as normative mappings relevant to understanding the philosophical complexity *and* ethical impact of this technology. Some things, like ontological indeterminacy, will be impossible to design for. Other things, we argue, can and should be considered.

This section will offer selected normative frameworks which aim to provide guidance. A normative framework is a model for approaching, in this case, XRT design which identifies and balances trade-offs between normative goals. The normative framework does not itself give clear answers to the question, but rather provides strategies for answering these questions. Frameworks described here will be 1) Ethics by Design, 2) Value-Sensitive Design, and 3) the Human Development Approach. An overview of the philosophical and ethical issues outlined in D2.1., D2.2., D2.3, and Part 1 of this document are available for reference in the Appendix.

However, before we can discuss the proposed normative framework, it is necessary to first answer the question, "do developers have a moral responsibility to develop ethically sensitive XRTs?" Note by developers, we refer to all stakeholders who are responsible for and involved in developing XRTs, including but not limited to technology creators and innovators as well as industry stakeholders.

The Moral Responsibility of Developers to Develop Ethically Sensitive XRTs

The discussion on the moral responsibility of developers to develop ethically sensitive XRTs is part of the larger philosophical discussions on what moral responsibility means. It is beyond the scope of this document to provide an overview of these discussions. Instead, we will condense aspects of moral responsibility that various approaches may have in common and briefly examine if and how this applies to developers. To do this, we looked at the philosophical approaches identified by Matthew Talbert on "Moral Responsibility" in the Stanford Encyclopedia of Philosophy and sifted the aspects of moral responsibility that are present on the various approaches.[23] (Talbert 2024) Equipped with the findings, we shall respond to the question whether developers have a moral responsibility to create ethically sensitive XRTs.

Moral responsibility exists when the following conditions are present in the agent:

[23] An interesting observation would be that the elements of moral responsibility outlined by Talbert align with the conditions of a moral act in classical Thomistic philosophy: (the object of the) act, intention, and circumstance. This observation may be some sort of a validation that contemporary reflections on moral responsibility coincide with Medieval reflections on the matter. Arguably, this is an attestation of the durability/robustness of the criteria mentioned above.

Funded by the European Union

- *Possession of capacity for agency (ability to understand the nature of their actions (including awareness of relevant moral norms and requirements as well as the ability to be sensitive to considerations that bear on their actions) and to exercise control over their behavior)*[1]. In the case of the XRT developer, this only requires having the faculty for choice and understanding -- including basic knowledge of right and wrong that in this case is associated with the act of developing a technology and the capacity to reflect on consequences -- mental faculties that the law (rightly or wrongly) usually associates with individuals within the age of majority/adults;
- *The action arising from the agent's own internal states (i.e., that the agent intends this action, with the understanding that intent to perform an act can have various degrees), rather than being externally imposed*[1]. In the case of the XRT developer, this means that the act of developing is not done under duress, manipulation, or coercion.

Note that moral responsibility may have degrees, i.e., that a person may be wholly or partially responsible for an action, or in interpersonal settings, that one person may have more responsibility than another. The degree of knowledge, intent, and will; the level of participation in the act; and the presence of extenuating circumstances, such as the nature of hierarchies and other interpersonal dimensions, all contribute to determining the extent of moral responsibility. The typical example in the literature would be cases involving command or hierarchical responsibility when examining or litigating war crimes. In the case of developing XRTs, a programmer involved in the development of a socially risky technology (think of the many technologies which have strong dual-use risks) whose family income depends on the job may still intend to develop this technology due to practical reasons, but this person's responsibility cannot be weaker than the CEO responsible for planning and financing this project.

Based on the discussion above, the basic presumption should be that developers are responsible for the technology that they develop, including of course all the intended and foreseeable features of this XRT. In this sense, developers are ethically required to ensure that their technology has a degree of ethical sensitivity, i.e., that right or wrong, good or bad, and considerations of human flourishing such as well-being, are considered in the planning, developing, and deployment of technology. What this means in terms of what must be considered will be the goal of the forthcoming sections.

Ethically sensitive XRT

In responding to the question of what it means for developers to develop ethically sensitive XRTs, utilizing perspectives that emerge in the current literature on reflections on emerging technologies, we propose to use as reflection tools three perspectives: Ethics by Design, Value-Sensitive Design, and the Human Development Approach. (van den Hoven 2012, Oosterlaken 2014, Davis and Nathan 2015, Epley and Tannenbaum 2017, Dainow and Brey 2021) We will first explore what these three perspectives are and then attempt at applying these perspectives on the question, "what is meant by an ethically sensitive XRT?"

Ethics by Design

There is widespread use of this perspective, with publications from the European Commission and the World Economic Forum, to name a few. (Dainow and Brey 2021, World Eocnomic Forum 2020) The core message is that ethics must be treated not simply as an "individual issue," but rather an issue of design, i.e., that agents ought to "create contexts that promote ethical actions." (Epley and Tannenbaum 2017) These agents must proactively create spaces that encourage ethical behavior

because context and environment play a crucial role in shaping behavior. Such a perspective is an offshoot of findings in behavioral psychology such as those by Richard Thaler and Cass Sunstein on nudges. (Thaler and Sunstein 2008) In the latest edition of their book, Thaler and Sunstein claim that respecting other's freedom should not be equated to simply maximization of choices (i.e., simply giving people as many choices as possible) because whoever is responsible for the architecture of any context[24] will inevitably influence the choice or behavior of (many/some) other people who enter or use the created space. (Thaler and Sustein 2022) Thaler and Sunstein call such agents "choice architects." (Ibid.)

This perspective is interesting as other readers might (mis)construe it as opposing moral responsibility. At least for the proponents of nudges and Ethics by Design, this is not exactly the case. Rather, they emphasize that ethical behavior is not solely a product of individual character, and that context and environment play a crucial role in shaping behavior. By focusing on design, agents can create systems that make it easier for individuals to act ethically. But we digress. Going back to Ethics by Design.

In Ethics by Design, three psychological processes are crucial for choice architects: attention, construal, and motivation. (Epley and Tannenbaum 2017)

- Attention is naturally limited, and as such, decision-making is usually limited to whichever information is readily available/accessible. This being the case, the choice architect must design their context in a manner that allows individuals who enter or use the context to put their attention to ethics to increase the likelihood of ethical behavior. (Ibid.) To quote Epley and Tannenbaum,

  > Effective ethical systems induce people to think about ethics routinely. Examples of triggers include ethics checklists filled out before making a decision, messages that make ethical principles salient in the environment, or heuristics that can become repeated mantras for ethical action. (Ibid.)

- Construal refers to how people interpret their environment, their reaction to the environment being a function of this construal. This means that a change in construal may or will likely affect how human beings will react. For choice architects, this means designing a context that encourage ethical construals. (Ibid.) Again citing Epley and Tannenbaum, "Inducing employees to ask themselves 'Is it right?' rather than 'Is it legal?' should lead to an increase in prosocial behavior." (Ibid.)
- Motivation is driven by incentives: we are usually motivated to do something when there is an incentive attached to the act. Findings in moral psychology have demonstrated that prosocial motives are actually powerful motivations for ethical action. (Ibid.) Thus, being presented with a video of a potential beneficiary who clearly needs aid usually gets more donations than charity drives that do not provide this much information. Along with the observation that morality is often socially influenced/malleable[25], Epley and Tannenbaum define an ethical

[24] We use architecture here not as the practice of designing and constructing buildings but to act of designing any context (whatever this might be, be it an environment, a setting, an artefact, a document, among others) by an agent – who Thaler and Sunstein call choice architects. From this perspective, I am the choice architect of my bedroom and someone from the government is the choice architect of tax forms, etc.

[25] For those interested in this topic, the Stanford Prison Experiment and the Milgram Experiment provide interesting results.

Funded by the European Union

context as one where “systems promoting ethical behaviour create opportunities for acts of goodwill and emphasize positive actions to establish more ethical norms.” (Ibid.)

Implications for XRT development

Incorporating the principles of Ethics by Design into the development of XRT can lead to several recommendations, such as the following:

- Ethics must be essentially incorporated into design from the start of the process. This means, among other things, the use of an ethics framework to guide the process of development.
- Leverage attention to foster ethical awareness. This could mean including reminders within the design of immersive environments, such as virtual meeting places, to prioritize respect and inclusivity, for example. The user interface (UI) can also be designed in such a way that ethical behavior is encouraged. For example, to make the principles of informed consent and transparency salient, when a user tries to record a session, the UI could display a notification/visual icon that indicates to everyone that the session is being recorded.
- Incorporate ethical construals in XR environment design. This may mean the use of contextual cues as nudges to users. For example, activating an AR camera in shared spaces would prompt a message stating, "This area may include individuals who have not consented to being recorded."
- Motivate ethical behavior using prosocial incentives, such as highlighting community members who contribute constructively to shared XR spaces.
- Establish clear community standards that encourage ethical interactions.
- Embed interactive ethical decision-making tools, for example, a sustainability checklist for MR applications.

Value-Sensitive Design

Value-Sensitive Design is a perspective that was pioneered by Batya Friedman from the University of Washington. Its approach is for developers to “proactively consider human values throughout the process of technology design.” (Davis and Nathan 2015) This springs from the assumption that technology is value-laden (as opposed to value-neutral), and that how a specific technological artefact is designed may promote or hinder values. As such, values must be purposely and proactively integrated in the design process, from the early stages to the final stage of technological development. The question arises, what is meant by “value”? In this approach, value is “what a person or group of people consider important in life” and further makes a distinction between moral and societal values. Moral values apply to human conduct and character necessary for human well-being while societal values refer to what a society might consider to be important for collective well-being. (Ibid.) Examples of moral values include safety, privacy, sustainability, accountability, autonomy, trust, justice, equality, responsibility, etc. Examples of societal values include democracy, global development, social improvement, cohesion, inclusion, environmental sustainability, etc. (Ibid.)

Specific to Value-Sensitive Design is the *interactional understanding of technological appropriation* which means that “values are not embedded within a technology; rather, they are implicated through engagement.” (Ibid.) Technological design promotes or hinders value depending on how it is appropriated by individuals and societies that live within a specific context. (Ibid.) For proponents of Value-Sensitive Design, therefore, the process of integrating values in technological design does not happen in a vacuum; rather, along with design should come considerations about context and people

Funded by the European Union

(both direct and indirect stakeholders). This makes the integration of values in the development process iterative and requires specific value "investigations," i.e., conceptual investigations, empirical investigations, and technical investigations. (Ibid.)

Conceptual investigations refer to two areas of investigation: identification of direct and indirect stakeholders who are likely to be affected by the technology; and identification of the values that are to be implicated when the direct stakeholders use the technology. In the case, for example, of a VR platform for educational purposes, the direct stakeholders would be the students, the teachers, and most probably the school administrators. Indirect stakeholders would be the parents of these children, and depending on whose vantage this investigation is made, probably also education policymakers. As to the values that are likely to be implicated, without consulting the stakeholders and probably through the developers' knowledge of the literature, these would likely be engagement, accessibility, privacy, safety, among other possibilities, with the possibility of a conflict between engagement and privacy/safety.

Empirical investigations require understanding of the stakeholders "understandings, contexts, and experiences", as well as their observable actions and their "concerns, reflections, and aspirations", relevant to the technology and the implicated values. (Ibid.) The potential value conflicts identified during the conceptual investigations may also be explored here. This phase is usually done through direct consultation with the stakeholders through various qualitative and quantitative methodologies including interviews, focus-group discussions, surveys, participant observation, and artifact analyses. In the case we provided above, this could mean several activities such as interviews with teachers on their perception of the platform's usability and fit with the curriculum; focus-group discussions with students to understand their views of the usability of the platform (knowing, for example, what they found useful, exciting, or frustrating); doing classroom observations to see how the XR application affects the class, i.e., whether it enhances or disrupts learning; among others.

Technical investigations to explore how the technology's features aid or hinder certain values. Studies on technological design in support of a specific value or analysis of how a feature implicate specific values fall within this set of investigations. (Ibid.) Going back to our case above and based on the inputs from the two sets of investigations, this could be analysis of how headset ergonomics can be improved to ensure the students' comfort, especially during long sessions; refining data collection process to minimize privacy risks; exploring how the XR interface may include accessibility features such as subtitles, among others.

These investigations are meant to be iterative and integrative, i.e., that the investigative process is repeated and refined continuously to ensure that the values are appropriately considered and incorporated throughout the development stages. (Ibid.)

Implications for XRT Development

In section 3.2., when discussing the three value investigations, we used the example of a XR platform for educational purposes, which illustrated the application of value-sensitive design in XRT development. Below are some recommendations on how Value-Sensitive Design may be applied in the development of XRTs.

- Developers should prioritize value-integration throughout XRT development by proactively incorporating moral and societal values in all stages of XRT development. With regard to societal values, a useful resource for developers would be the resulting documents from the European Citizens' Virtual Worlds Panel. (European Commission 2024)

- Developers must ensure that design is stakeholder centered. This means that both direct and indirect stakeholders are properly identified and that the features of XRTs address the diverse needs, aspirations, and concerns of these groups.
- Developers must be sensitive to contextual variations: the implications of XRTs vary across contexts, and as such, design decisions should, as much as possible, be respectful of the various contexts (cultural, social environmental) where the technology will be used.
- Value conflicts must be anticipated and identified early in the design process and ideally resolved during the developmental process. Use iterative feedback loops and stakeholder consultation methods in identifying and resolving these conflicts.
- Developers should use iterative and integrative processes. This means that XRT development must be treated as a continuous cycle of conceptual, empirical, and technical investigations.
- Anticipate, and as much as possible, address potential long-term impacts, such as socio-cultural and environmental implications of XRTs over time.

Human Development Approach

The Human Development Approach, also known as the Capabilities Approach, is a conceptual framework pioneered by the economist Amartya Sen and the philosopher Martha Nussbaum which has been utilized in various areas and sectors including but not limited to economics, politics, public health, global health, and, of course and more recently, in technological development. Within these areas, the approach has been used both in the planning and/or evaluation of regulations, guidelines, programs, interventions, social arrangements, and/or well-being. Basic to this approach is the interest on "justice, equality, well-being and development," (Oosterlaken 2012) or, as UNDP defines it,

> Human development – or the human development approach – is about expanding the richness of human life, rather than simply the richness of the economy in which human beings live. It is an approach that is focused on people and their opportunities and choices.

The unit of analysis (or the "evaluative focus", as Sen puts it) of this approach "can be either on the realized functionings (what a person is actually able to do) or on the capability set of alternatives she has (her real opportunities)." (Sen 1999) According to Robeyns, the difference between functionings and capabilities lies in what is realized versus what is effectively possible. (Oosterlaken 2012) To provide an example, functionings would be various states of being and doing such as resting, walking, being healthy, being a student, etc. Capabilities, on the other hand, refers to the real possibility to rest, walk, be healthy, be a student, etc. One's "capability set" would then refer to "the alternative functioning vectors" that a person may choose from and thus, a representation of this person's freedom. (Ibid.)

Before we explicate the relevance of technology in this discussion on capabilities and functionings, it would be worthwhile to point out that though Sen and Nussbaum largely agree on the goal of the approach as well as on the distinction between capabilities and functionings, they differ on the opinion of creating a list of capabilities. Nussbaum has such as an "essential" list which she calls "Central human capabilities". These are "central" capabilities in the sense that each and every item on this list acts as a threshold, an "account of minimum core social entitlements" for a life worthy of human dignity. (Nussbaum 2006) The list consists of the following capabilities: life; bodily health; bodily integrity; senses, imagination and thought; emotions; practical reason; affiliation; other species; play; control over one's environment. (Ibid.) We will go back to this list later in this section.

Funded by the European Union

Within this approach, there were several reflections on the place of technology and the aim of technological design and development. Van den Hoven, for example, makes the observation about the intimate connection between humans, technology, and technological development because several human capabilities only are through technology. (van den Hoven 2012) Humans have an innate need for (making/creating) complementary objects that make us *capable* of protecting us, making our lives easier, and/or making our lives pleasurable. For example, we cannot protect ourselves from the weather without clothes or infrastructure. We cannot write anything without a pen (or, nowadays, a computer). We cannot travel intercontinentally without the appropriate mode of transportation. The list goes on. In this sense, van der Hoven considers technologies as "agentive amplifiers." (Ibid.)

The fact that technologies can be agentive amplifiers does not mean that all technologies actually are. There is a specific sequence when we talk about the conversion of a technology into a human capability. The sequence usually goes as follows: the object (a pen) to the object's characteristics (a writing instrument) to the capability to write, to the functioning (or act) or writing. (Ibid.)Within this sequence, there could be several factors, including technological design, or other personal, social, economic, and other such factors that could hinder the conversion of an object to capabilities. Such factors could be a defective pen, an inappropriately sized pen, my inability to use a pen, my inability to write, etc. In identifying attenuating factors to the conversion, some clearly pertain to design.

Given that technological design and development can either amplify or attenuate human capabilities, scholars such as Oosterlaken argue that "design should aim to enhance people's real opportunities to do and be what they value" (Oosterlaken 2012) Some scholars, such as Zheng and Stahl, even created a framework partly based on the Human Development Approach that may be used in the development and evaluation of technologies. (Zheng and Stahl 2012) This framework has four central points, which can be expanded to create a checklist:

- Human-centered technological development: technologies are meant to serve the expansion of human freedom and opportunities, allowing humans to lead lives they value.
- Human diversity: consideration should be given to the diversity of human conditions and the factors, including conversion factors, that support or attenuate human capabilities/capability sets.
- Protecting human diversity: technology should not diminish people's ability to act and make choices.
- Democratic discourse: public dialogue and inclusion in the development of a technology as a necessity.

When reflecting on the first three points, Nussbaum's list of central capabilities could come in handy.

Implications for XRT Development

According to the Human Development Approach, XRT development must aim to support the enhancement of human capabilities and the promotion of opportunities for meaningful actions and choices. Below are a few recommendations:

- XRTs should aid in the expansion of human capabilities; conversely, XRTs should minimize attenuating human capabilities. This means, for example, that addictive game mechanics and other potentially exploitative traits must be addressed.
- XRT development should take into account human diversity. Concretely, this means that developers should recognize barriers that may hinder access. Examples would be addressing the adjustability of the various parts of VR headsets to accommodate physical differences in

Funded by the European Union

head sizes, head shapes, pupillary distance, eye condition, and/or other relevant motor disabilities. There should also be considerations for language inclusivity through, for example, multilingual support.

- XRT development must include democratic discourse through means such as participatory workshops, public consultation, and/or the inclusion of citizen science approaches.
- Promote equitable access to XRT technologies by, for example, making XRTs that are affordable and accessible even to regions with limited access to the internet.

Conclusion: The Responsibility of Developers in Creating Ethically Sensitive XRTs

The development of XRTs presents opportunities for advancing human capabilities, fostering inclusivity, and promoting well-being. It is therefore critical to gain a thorough understanding of how XR relates to and may uniquely impact persons – at the individual, inter-personal and sociological levels. This requires analyses in terms of impacts on health and well-being but also philosophical and normative impacts. As such, Part 1 of this paper iterated the philosophical and normative dimensions of XR technologies which could not be conceptualized in terms of risks or harms. We identified several issues at the level of ontology, epistemology and ethics, including (but not limited to) whether immersive environments are real, impacts on identity, difficulties in ascertaining knowledge and making judgements (both general and moral), and problems related to meaningfulness. In Part 2, we argued that there must be real intent and resolve to integrate ethics in the development process. Reflecting on the perspectives of Ethics by Design, Value-Sensitive Design, and the Human Development Approach, the following key responsibilities emerge: ethics must be embedded throughout the design process; human diversity and accessibility must be acknowledged as important considerations in XRT development; XRT design must purposely foster empowerment and prohibit exploitation; value conflicts must be anticipated and resolved; direct and indirect stakeholders must be included in the development process; XRT development should be facilitative of, or at least not attenuative of, short-, mid-, long-term collective well-being.

Funded by the European Union

References

Dainow, B. & Brey, P. *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (2021).

Davis, J. & Nathan, L. Value Sensitive Design: Applications, Adaptations, and Critiques. in *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (eds. van den Hoven, J., Vermaas, P. & van de Poel, I.) (Springer Science + Business Media, Dordrecht, 2015).

Epley, N. & Tannenbaum, D. Treating ethics as a design problem. *Behavioral Science & Policy* **3**, 73–84 (2017).

Ethics by Design: An organizational approach to responsible use of technology | World Economic Forum. https://www.weforum.org/publications/ethics-by-design-an-organizational-approach-to-responsible-use-of-technology/.

Nussbaum, M. C. *Frontiers of Justice: Disability, Nationality, Species Membership*. (Harvard University, Cambridge, 2006).

Oosterlaken, I. Human Capabilities in Design for Values. in *Handbook of Ethics, Values, and Technological Design* (2014). doi:10.1007/978-94-007-6994-6_7-1.

Oosterlaken, I. The Capability Approach, Technology and Design: Taking Stock and Looking Ahead. in *The Capability Approach, Technology and Design* (eds. Oosterlaken, I. & van den Hoven, J.) 3–26 (Springer Science + Business Media, Dordrecht, 2012). doi:10.1007/978-94-007-3879-9_1.

Sen, A. *Development as Freedom*. (Anchor Books, New York, 1999).

Talbert, M. Moral Responsibility. *Stanford Encyclopedia of Philosophy* (2024).

Thaler, R. & Sunstein, C. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. (Yale University Press, 2008).

Thaler, R. & Sunstein, C. *Nudge: The Final Edition*. (Penguin Random House, Dublin, 2022).

van den Hoven, J. Human Capabilities and Technology. in *The Capability Approach, Technology 27 and Design* (eds. Oosterlaken, I. & van den Hoven, J.) (Springer Science + Business Media, Dordrecht, 2012).

Virtual worlds panel - European Commission. https://citizens.ec.europa.eu/virtual-worlds-panel_en.

Zheng, Y. & Stahl, C. Evaluating Emerging ICTs: A Critical Capability Approach of Technology. in *The Capability Approach, Technology and Design* (eds. Oosterlaken, I. & van den Hoven, J.) (Springer Science + Business Media, Dordrecht, 2012).

APPENDIX

Table 1. ETHICAL & PHILOSOPHICAL ISSUES IN THE USE OF XR TECHNOLOGIES		
1. Ontological and Epistemological Issues	Reality and Virtuality	XR raises questions about whether virtual objects and environments can be considered "real." This distinction affects how users perceive and interact with virtual entities. Virtual spaces blur the boundaries between digital and physical realities, leading to discussions about the nature of existence and the meaning of presence in XR.
	Knowledge Formation	Users often depend on visual, auditory, and haptic feedback in XR environments to form their beliefs about the surroundings. However, these sensory inputs can be altered or manipulated, complicating the process of discerning what is true or authentic. Epistemic uncertainty, such as differentiating between virtual and physical interactions, poses significant challenges to traditional methods of reasoning and decision-making.
	Causality and Interaction	Actions in XR may lack the direct physical consequences associated with real-world actions but still have meaningful impacts on users' perceptions and experiences. Ethical concerns arise when interactions feel "real" but lack physical world causality.
2. Identity and Representation	Digital Identity	XR enables users to create digital personas that may differ significantly from their physical identities. This raises concerns about authenticity, manipulation, and control over these virtual representations. Identity theft and unauthorized avatar manipulation can erode trust and compromise user autonomy.
	Self and Social Perception	XR reshapes how individuals perceive themselves and others, potentially influencing real-world self-esteem, body image, and interpersonal relationships. Social dynamics in virtual spaces may reflect or reinforce biases and stereotypes from physical reality, necessitating inclusive design approaches.
	Inclusivity in Representation	Ethical design of avatars and XR systems must account for diverse cultural, gender, and disability perspectives to prevent exclusion or misrepresentation of marginalized groups.
3. Moral and Ethical Judgment	Ethics in Virtual Environments	Virtual environments often replicate real-world scenarios, allowing users to engage in morally ambiguous or harmful behaviors without direct consequences. This raises questions about whether these actions are ethically permissible. Harassment, exploitation, and violence in XR are particularly concerning because they can evoke psychological harm even in virtual contexts.

Funded by the European Union

	Ethical Boundaries	XR challenges conventional ethical boundaries by enabling actions that are impossible in physical reality. For instance, the ability to harm virtual representations of others requires rethinking accountability and moral responsibility
4. Autonomy and Privacy	User autonomy	XR experiences may subtly influence users' decisions through immersive design, raising concerns about manipulation and the erosion of free will. Ensuring users have control over their virtual experiences, from interactions to data collection, is essential for preserving autonomy.
	Data privacy	XR systems collect extensive biometric and behavioral data, which, if misused, can violate users' privacy and expose them to risks such as identity theft.
	Surveillance and consent	The immersive nature of XR makes users more vulnerable to pervasive surveillance. Clear and transparent consent mechanisms are critical to mitigate this risk.
5. Social and Structural Impacts	Communication and Social Norms	XR has the potential to fundamentally reshape how people communicate, both in digital and physical spaces. This includes creating new norms for behavior and etiquette in virtual environments. Virtual interactions may reduce face-to-face engagement, raising concerns about their effects on societal cohesion and empathy.
	Inequalities in Access	High costs, technical barriers, and lack of infrastructure can prevent economically disadvantaged groups from accessing XR technologies, exacerbating the digital divide.
	Economic and Employment Effects	XR could disrupt traditional labor markets by automating roles or creating entirely new industries, requiring proactive measures to ensure equitable opportunities.
6. Sustainability and Environmental Ethics	Environmental Costs	XR technologies are resource-intensive, with significant energy demands for data processing, storage, and transmission. These systems contribute to carbon emissions, necessitating sustainable design practices.
	Long-Term Implications	The cumulative environmental and societal impacts of widespread XR adoption remain poorly understood, highlighting the need for ongoing research and proactive mitigation.
7. Normative and Foundational Principles	Justice and Equity	Ethical frameworks for XR must address systemic biases to ensure all users, regardless of socioeconomic status or identity, can benefit equitably from these technologies.
	Transparency and Trust	Users need clarity on how XR systems operate, including the handling of their data and the functioning of algorithms, to build trust and foster informed consent.
	Design for Human Flourishing	Developers should prioritize creating XR experiences that promote positive outcomes,

		such as psychological well-being, inclusivity, and meaningful social engagement.
Philosophical considerations	Meaning-Making and Identity	XR acts as a medium for shaping personal and collective identities, influencing how users perceive themselves and their place in society. Virtual environments serve as meaning-making spaces, offering opportunities for self-expression but also raising concerns about dependency and escapism.
	Mediation of Reality	By mediating how users perceive and interact within virtual worlds, XR technologies may play a powerful role in shaping beliefs and societal norms external to those worlds, warranting careful oversight.
Long-Term Risks and Uncertainties	Emergent Risks	As XR technologies evolve, unforeseen risks may emerge, such as new psychological or societal impacts.
	Knowledge Gaps	Research is still lacking on the long-term effects of XR on individuals and societies, particularly regarding cumulative exposure and generational adoption.

Funded by the European Union