In Focus

Virtual Representations and Their Ethical Implications

sciendo

ROBERT MCNAMARA, Tallinn University, Estonia; e-mail: robertmc@tlu.ee

10.2478/bsmr-2023-0011

ABSTRACT

This paper will address ethical concerns surrounding the representation of vulnerable groups as well as the methodological challenges inherent in using artificial intelligence and human-like computer-generated characters in human studies that involve representing such groups. Such concerns focus on consequences arising from the technological affordances of new systems for creating narratives, as well as graphical and audio representations that are capable of portraying beings with close resemblance to humans. Enacting such virtual representations of humans inevitably gives rise to important ethical questions: (1) Who has the right to tell certain stories? (2) Is it ethical to change the medium of a narrative and the identity of a protagonist? (3) Do such changes, or technological mediations, affect whether a vulnerable group will be fairly and accurately portrayed? (4) And what are the implications, either way? While the backdrop of the paper involves discussing the potential of virtual representation as a meditative tool for moral and social change, the ethical implications inherent in the use of new cutting-edge technologies, such as OpenAI's ChatGPT and Unreal Engine's MetaHuman, to create human-like virtual character narratives call for theoretical scrutiny from a methodological perspective.

KEYWORDS: technological mediation, ethics, ChatGPT, tools of virtuality

INTRODUCTION

The general focus of current research has been to study how tools of virtuality, as artistic tools, might mediate narrative understanding and what effect this might have on decision-making or emotionally motivated behavior and outlook. Particular tools of virtuality under discussion consist of artificial intelligence (AI) generative models, as well as more or less lifelike computer-generated characters (CGCs) used for the purposes of representation and authorship in storytelling. The degree to which virtual representation or authorship might be considered to depict an out-group perhaps determines the ethics of such authorship, or at least starts to raise questions of ethics based on relationality. Considering the broad spectrum of the topic, this article will focus primarily on refugees as the paragon of out-group vs. in-group membership. This concept, however, could also be extrapolated to other groups who are marginalized or face similar biases

leading to exclusion. For the purposes of this discussion, human-created virtual representation includes AI representation, in the sense that AI representation can be thought of as not only a tool created by and used to generate narratives but also something that inherently reflects the human world back upon itself. Another interesting aspect of this representative world generated by interpolating datasets is that such datasets inherently represent only certain anthropological or cultural artifacts, while excluding other human cultural sources and narratives. While AI may at first blush appear to be an independent entity, it cannot be separated from the datasets that derive from human experience and artifacts. In designing academic or scientific studies involving tools of virtuality, or in artistic experimentation with technological mediation to design and study narratives or their effect, researchers will inevitably confront ethical questions that may have no clear or simple answer.

In Charlie Brooker's latest season of *Black Mirror*, he presages a dystopian future for art, in the episode "Joan Is Awful," wherein a quantum computer employing AI exercises its capacity for recording events from a person's smartphone camera while near-simultaneously producing a lifelike computer graphics-based digital likeness of an actor portraying the recorded person in a computer-generated narrative film of their life as recorded, and with poetic license to alter certain facts (Brooker, Pankiw 2023). Although it is clearly a work of fiction, Brooker delves into a technological moment toward which society may be clearly headed. For instance, as this is being written, Hollywood writers and actors are currently on strike, not necessarily with a primary motivation for more money but with a mind toward survival: their key contract term that is nonnegotiable is to keep AI out of scriptwriting.

Examining the current scriptwriting capabilities of OpenAI's ChatGPT, as well as the animated human-likenesses of CGCs from Unreal Engine's MetaHuman and other production technologies such as deep fakes, it becomes clear that not just scriptwriters but also actors, directors, as well as much of the current paradigmatic scaffolding of what is thought to be the primary filmmaking (as well as artistic) methodology appear to be in for a sea change. Already, perhaps cast into the same principled level of resistance as Luddites, those who manage and benefit from keeping close control over what should be seen as art are already mobilizing if not for an ideological defense of what the idea of "true" art should be then in support of the entrenched interests of traditional artists to continue to make a living from art, in the way they believe "art should be defined."

Yet, as we may be getting closer to actual quantum computing, as central processing units (CPUs) continue to relentlessly chase a runaway Moore's Law, what we think of as art is being challenged and redefined, despite hegemonic feelings or traditions. Art mediates humans much in the way that technology does, and so the inevitability of a marriage of the two will continue, as has always been the case. The artist, the context, the artistic tool and medium, just as the world itself, as experienced, grow out of the same spectral moment to form new representations that are constantly redefining art. From the *techne* of crushing natural materials down into inks or paints in an effort to represent the world as symbols on the inside of caves, to a time even long before language, expression and its tools have been a constant human occupation. Resisting technologically mediated art creation through AI or CGCs, however, is as futile as a painter at the dawn of the computer age protesting against graphic design as *not* art. And yet, continue to protest we do. For it is not only scriptwriters that are at risk of being put out of work; a litany of jobs and professions may be affected, for instance graphic designers, coders, literati, (language) teachers, content creators, advertisers, researchers, analysts, historians, and the list goes on. The "art" of several professions may soon become transformed; with this trans-

formation, new forms of authorship and representation will and have emerged.

Those who identify as artists in particular are perceived to be in the crosshairs of new technologies, whether this is actually true or not. For instance, in a September 2022 *Washington Post* article entitled "He Used AI to Win a Fine-Arts Competition. Was It Cheating?," the reporter Drew Harwell chronicled the controversy of game-designer Jason Allen's AI piece, entitled "Théâtre D'opéra Spatial," which he created using an AI-based application called Midjourney, and which won first prize, beating out several other competitors at the Colorado State Fair in the "digitally manipulated photography category" (Harwell 2022). While Allen expressed to the paper that he had received extensive criticism online for his creation – including a tweet expressing that his work marked "'the death of artistry'" – a representative for the agency responsible for the fair noted that "'People put bananas on the wall and called it art, ... Even photography was not considered an art form for a long time; people said it was just pushing a button, and now we realize it's about composition, color, light. Who are we to say that AI is not the same way?'" (ibid.). But is AI really the same way?

Because the subject of representation in this instance involves an abstract, unidentifiable, or anonymous subject or group, seemingly outside of history, in an unknown future, the issue of authorship is not necessarily controversial due to issues of representation. Were the subject matter portrayed to be grounded in a reality involving a specific group or culture, then the author's identity might take on more relevance in terms of ethics of representation. Despite the defense of AI as an artistic tool analogous to a camera, it seems that much of the discontent with tools of virtuality, especially pertaining to those that cull from larger datasets or databases for their raw training data, stems from the fact that the imagery that is being included in the algorithmic recombination by machine learning (ML) comes from the works of others, although the recombinant nature of such AI models may in some cases be modeling from literally billions of images. When asked for comment on the issue of using AI to win the state fair, a local oil painter, Gregory Block, noted that what is art is



FIGURE 1. Théâtre D'opéra Spatial by Jason Allen in Harwell 2022.

essentially in the eye of the beholder. He notes that art

> 'doesn't have to necessarily be created out of a human soul, the artwork itself. It is for us to see and react to ... We the viewers are, in the end, the ultimate artists. We're the ones creating the world that is coming in through our eyeballs. That world is in our mind.' (ibid.)

In terms of AI, however, the question of whose "mind" the artistically represented world is "in" is a particularly thorny one. AI does not have a mind of its own. So whose mind is it? Buolamwini and Gebru (2018) point to the particular problem that AI datasets, from which such algorithms call forth such representations, are necessarily historically biased, as the well of artistic content from which they draw is over-represented by contributions from certain groups or cultures, while other groups of cultures may be completely absent. While viewers of art may be "in the end, the ultimate artists," such a view also neglects other ethical issues that may arise in the creation of art through tools of virtuality, which might not be so easily answered with the art viewer taking on ultimate responsibility for the interpretation of an artistic message. For instance, in Allen's case, an unidentified Victorian woman is depicted with no face, in a purported futuristic setting; but what about art that conveys the narrative of someone more identifiable, what about art that seeks to tell the story of vulnerable or other excluded groups? While newfound tools of technological mediation may exist for those willing to explore their potential, do we have the right to use them in whatever way we wish?

The debate at hand regarding tools of virtuality and artistic mediation through the use of virtual tools revolves not just around how art *should* be made; it also gives rise to other, perhaps deeper questions, such as:

1. Who has the right to tell certain stories?
2. Is it ethical to change the medium of a narrative and the identity of a protagonist?
3. Do such changes, or technological mediations, affect whether a vulnerable group will be fairly and accurately portrayed?
4. And what are the implications, either way?

Enacting such virtual representations of other humans and their stories inevitably creates ethical questions, especially when moving from pictorial representation to narrative depiction, although images themselves may carry many of the same questions. Throughout the following paper I will try to outline what my current research into AI and CGCs involves by way of illustration and as a backdrop for some of the theoretical concepts and related ethical implications of using such technology to artistically depict real-life scenarios, particularly in the context of representing those that might be considered vulnerable groups.

WHO HAS THE RIGHT TO TELL CERTAIN STORIES?

In general, from the perspective that art is for the interpretation of the reader or enjoyer of the art, and that the creator or producer is somehow irrelevant or less important, it would perhaps not matter who tells the story; however, this calls up the very real problem of who has the right to tell which narrative. In David MacDougal's book *Transcultural Cinema*, in the chapter entitled "Whose story is it?," he notes that the subject of ethnographic films has taken on new importance, with a shifted emphasis from the voice or point of view of the film's author (1998: 150). He focuses on subjectivity, rather than objectivity, as a means of creating a more inclusive artform through mutual artistic or shared work (ibid.). This form of filmmaking has been referenced as "participatory" (Grimshaw 2005: 25) in the sense that the creator and the subject both engage in the "social expe-

rience" (MacDougal 2001: 71) of the artistic production, and this relationship is made more explicit as a "dialogue or a conversation between different parties, rather than a task animated by notions of data collection and scientific interpretation" (Grimshaw 2005: 25).

In the case of AI-mediated storytelling or narratives, this aspect seems somewhat diluted, in the sense that there is not necessarily explicit participation in a social experience, but rather an amalgamation of the experiences and narratives contained in AI training datasets referencing sometimes many millions of individuals. This creates quite a different dynamic by limiting individual representation, while at the same time encompassing an expression that was previously impossible to convey, recombinant amalgamation of culture as seen through the prism of vast data analysis. In the case of my research, the question of whether this amalgamation is a fair and accurate representation arises because, through the use of tools of virtuality, facts of an individual refugee's narrative are mediated by AI from interpolation of an expansive language dataset, ChatGPT. True facts that were conveyed by an actual asylum seeker to a Finnish Immigration officer, and summarized in an academic journal, were mediated through multiple levels of AI and through the mouth of a CGC into a dramatic narrative for the purpose of studying empathy (see Bodström 2020: 638–641). Perhaps similar to the depiction of an unwitting life of another as in *Black Mirror*, through tools of virtuality the story of an actual refugee is being told, and viewer reaction is being measured.

This begs many questions: Does this nameless (or pseudonym-protected) refugee, as mediated into an AI-amalgamation, then truly have a voice? What gives one the right to tell another's story? And, once amalgamated, who is this other? Does the telling of the story of another by yet another, or by culling through technologically mediated datasets to create a narrative entity, take away the original experiencer's voice, or might it translate the experience into a more universal language? Does it remove this person's ability to speak, even if perhaps this person can or will not? Could this be seen as a new interpretation, or as some sort of universal or archetypal representation beyond any given individual story or its teller? Should such artistic license be allowed? And, particularly, should it be allowed when dealing with the stories of vulnerable or excluded groups or populations, such as refugees, as we have here; especially when the datasets are incomplete and also contain exclusions themselves? Interestingly, such questions move us from questions of whether the tools of making art through virtuality are fair and acceptable ("Was it cheating?") to a larger question of intentionality when making an artistic narrative, whether mediated by technology or not.

Intentionality seems of particular importance when seeking to represent vulnerable or "subaltern" individuals or groups just as participatory filmmaking interaction seeks to be more of an inclusive attempt at representation. As mentioned, intentionality, however, may not be able to give validity to representation as implicit biases will potentially inevitably tend to skew communication. It is thought by some researchers that using AI might in some cases create more objectivity, and reduce human biases, in contradiction to the subjective approach of MacDougal (see O'Shea et al. 2018).

The idea of the "subaltern" originated from the existence of subordinates in the military, from which Antonio Gramsci borrowed and transformed the concept into that of the subaltern classes, as those who are excluded from decision-making processes; those who were and are subject to the hegemonic rule of superiors who set the tone, the rules, and the order of things (Green 2002: 1–2). At the periphery of the subaltern classes are thought to be those so excluded that, as Gramsci notes, they "'have not attained a consciousness of the class per se and … consequently do not even suspect that their history might possibly have any importance or that it might be of any value to leave documentary

evidence of it'" (Gramsci 1930: §48, as quoted in Green 2002: 11). This exclusion might be seen as analogous to Buolamwini and Gebru's (2018) assertion that certain groups are heavily underrepresented in computer datasets used in data science or to train ML language models.

Such a view would tend to support the idea of telling another's story; however, doing so may at the same time create a caricature of "the subaltern" that might thereby lead to negative outcomes, despite good intentions (see Morris, Spivak 2010). This might particularly be the case in employing AI, due to its tendency to amalgamate identities. The idea of the subaltern having a voice or finding representation within structures of colonial power may be essentially futile; the attempts of academics to help represent subaltern classes may lead to universalizing minorities into some reified entity (ibid.). Focusing on the stories of particularized individuals, however, or those real people being excluded from society, as individual human beings, should be examined as a worthwhile endeavor, especially if such people might never have had a voice. This intention might seem justified, with the caveat for vigilance in seeking reflexivity against reification (at least, in the pejorative sense). To apply Spivak's thinking to certain excluded people in populations such as refugees – who in many instances are not equipped to speak for themselves when it comes to the requirements of what James C. Scott mentions in his book *The Art of Not Being Governed* as legibility (Scott 2009) – it becomes clear that care should be taken as not only may this group often take on the role of the subaltern but also their role or identity necessarily may always be subject to being defined from the outside.

The ethical concerns centering on the potential problems with representations of historically excluded groups may go well beyond the question of whether they will be fairly and accurately portrayed. This is the case especially when the general portrayal is essentially defined as an outsider or out-group member. "Typically people and groups[] who are considered incapable of protecting their own interests[] are categorized as vulnerable" (Tiidenberg 2019: 4); however, they may not always see themselves that way. Refugees and asylum seekers generally fall squarely within this category, however, as the legal definition for determination of refugee status necessitates a legitimate fear of persecution, or vulnerability, and the very identities of citizen or asylum seeker are outwardly defined categories imposed on individuals. When doing research involving such populations, a project should attempt to "give marginalized populations a way to control the knowledge produced about them, as well as attempting to offer something of use or benefit to the participants or their communities" (Swartz 2011, as cited in Tiidenberg 2019: 5). This again returns to the idea of artistic intention, rather than tools.

In artistic or study design, when dealing with those that might be defined as excluded or vulnerable, artists and researchers should "'address specific sources of vulnerability in minority populations, rather than taking the vulnerability of minority groups to be homogenous'" (Rogers, Lange 2013: 2144, as cited in Hoffman, Jonas 2017: 6–7). Refugee needs and protections, however, will necessarily be balanced against the "social benefits" (Markham and Buchanan 2012: 4), the artistic aims, and the "moral imperative of research" (Hoffman, Jonas 2017: 10). Importantly, artistic representations, or representations made during research, ought to pay particular attention not to "reinforce and amplify existing hierarchies and biases" (ibid.: 14).

These types of concern and question tie into bigger issues regarding how populations are educated to understand "out-groups" (which is outside of the scope of this paper), which has strong bearing on the larger field of empathy literacy. Lacking empathy toward historically excluded groups, or having the capacity to turn it off, may be a common denominator to many current social problems. While homogenous representations of the refugee experience,

or reifications thereof, must be scrutinized, some narrativization mediated through AI language models might serve a methodological intent of translating the form in which the historically excluded individual's story is expressed. For instance, some studies have shown that AI models such as ChatGPT are more successful in communicating in a more empathetic way than human counterparts (Ayers et al. 2023) and exhibiting higher emotional awareness in blind comparisons to humans (Elyoseph et al. 2023). This may come down to expectations in communications, which language models appear to be particularly good at. In this way, language models might show affordances for giving voice to the narratives of the excluded, while also inevitably transforming them into more "legible" forms. As a technology with the potential capability of mediating social construction, or helping to allow its experiencers to perceive relationality, rather than essences, tools of virtuality and virtual narratives should be studied for their potentially important capacity for change.

IS IT ETHICAL TO CHANGE THE MEDIUM OF A NARRATIVE AND THE IDENTITY OF A PROTAGONIST?

As noted, the aim of my current artistic research project is to study whether an AI-mediated narrative and AI-mediated speech as told by various human-like CGCs have the ability to elicit empathy in the context of promoting social justice initiatives. The study recorded participants' empathy predisposition as measured by the Interpersonal Reactivity Index (see Davis 1983), which is a psychological instrument commonly used to measure empathy predisposition. Less than 2 percent of participants reported having an undocumented immigrant status. Participants were asked to adopt the role of an immigration officer tasked with making decisions about acceptance or rejection based on the narrative conveyed by the CGC seeking asylum. In this study a real legal standard was provided to participants, and they had to decide whether they believed the virtual asylum seeker had conveyed a "well-founded fear of persecution" (UNHCR 2019) based on religious belief (which could qualify for refugee status), or whether the CGC narrator was kidnapped simply because the kidnappers wanted money (which would not qualify for refugee status). In the study, the stories were offered to the participants by one of two different avatars; however, the narratives were the same. The difference between the avatars was in the degree of human-like realism, to determine whether the uncanniness of avatars has an effect on empathic decision-making. This was controlled against a text-based version of the same narrative.

A simplified description of the study hypothesis was that the more uncanny an avatar is, the more it might evoke cognitive empathy through the process of what has been called "distancing" and "defamiliarization" (Timmermans, Tavory 2012). Cognitive empathy is hypothesized to not only "compensate" for bias but also possess the capacity to be learned (see Di Palma et al. 2019). Whether low-empathy-predisposed participants exhibit a difference in decision-making, or demonstrate improvement in empathic behavior, based on avatar representation has profound implications for the increasing integration of virtuality and virtual characters across various future mediascapes. Currently, neurocinematic literature leaves an open question as to whether animated characters might evoke less cognitive empathy. For instance, "[a]ccording to a number of neuroimaging studies, the functional neural networks that are related to cognitive empathy, as well as related social perception and mentalizing, may not be as strongly activated by understanding a depicted character's 'intentional stance' from movement as compared to live-action characters (Tikka *et al*. 2021, citing Mar et al. 2007: 199; Han et al. 2005; Damiano, Dumouchel 2018). While these studies were not specifically focusing on correlations to cognitive empathy, they did examine cinema-evoked activation in the brain in terms of understanding a depicted

character's "intentional stance" from movement, which involved similar areas of the brain as involved with cognitive empathy. In general, whether evocation of cognitive empathy is changed by depictions of CGCs is relevant to and could have vast ramifications for what form machine-mediated virtuality takes and how it is employed in human interaction in digital and real-world spheres. Study results may also give insight into using avatars specifically as tools for moral and social change.

Interestingly, the outcome of the study seemed to demonstrate that the representative difference in uncanniness of the avatars was not significant; rather, the AI mediation of the narrative appeared to lead to widespread acceptance of both virtual avatars as well as the visually unrepresented asylum seeker based on the AI-mediated script alone. This outcome contrasted to the real-life outcome where the same core facts conveyed in a real-world scenario led to the actual denial of refugee status to the asylum seeker attempting to seek refuge in Finland. This result is in line with other studies that have looked at empathy and emotional affordances of ChatGPT to communicate, ironically, in a more human-like way (see, e.g., Ayers et al. 2023; Elyoseph et al. 2023).

In this case the medium of the narrative and the identity and narrative of the protagonist were changed, as the medium is portrayed in a computer-generated setting and the accuracy of the identity of the protagonist and his original narrative is unknown. While the facts of the narrative come from a published account of the actual refugee during a refugee status determination, the narrative itself was produced by an ML model. The facts of the narrative, from Erna Bodström's article "Asylum Decisions as a Performance: Intertextuality in Internal Credibility Assessment," outlined what the refugee said who was denied asylum by the Finnish Immigration Service, as the Service interpreted his kidnapping as motivated by monetary gain, rather than because he was a religious minority. The facts of his story were broken down into thirteen sentences and inputted into Open AI's ChatGPT model with the prompt to make a chronological first-person narrative based on the listed facts. The model then outputted the facts as a first-person story. The model was then prompted to make the story more "complex."

ML was used to construct the narrative rather than attempting to interpret or express the facts as an individual author, as, perhaps counterintuitively, the AI model, being based on training data from millions of sources, including the entirety of the Internet itself, was thought to be more objective and complete in its representation, as opposed to the narrative being interpreted from one subjective human standpoint. Using such language models also may serve to avoid individual biases (both in constructing narratives and in assessing them) as it may expand the potential for the narrative beyond what one person from one particular niche culture might believe an asylum seeker might talk like, how they might emote, or articulate to an immigration officer due to the language model's ability to amalgamate responses into a more emotionally appealing or cosmopolitan form. Using AI has been justified by government actors for similar purposes of eliminating bias and improving objectivity, even in the realm of refugee status determination and assessment (Tangermann 2017; Keung 2017; O'Shea et al. 2018; McNamara, Tikka 2023). As will be discussed, however, this methodology also suffers from potentially different ethical problems pertaining to the completeness and the accuracy of the training data used in the model.

After the ML language model created the script for the virtual asylum seeker, the script was plugged into another AI-based service, Murf.ai, which converted the text to speech with chosen parameters, using an artificial human-like voice. Another member of the Enactive Virtuality team at Tallinn University used an AI-powered application called Audio2Face to sync the speech file with the artificial human avatar, using Unreal Engine's MetaHuman. So in this

case, the narrative, originally drawn from a real case as portrayed in Bodström (2020), was spun through four layers of AI mediation and combined with two versions of a MetaHuman for inclusion in the study.

It is thought that this introduction of a narrative transformed from actual facts to a machine-mediated model of a persecution story allows the story to take on more universal narrative elements as the language model scrapes the entire Internet and beyond for collocations of accepted narrative universals. While this might be seen as a form of reification, it remains focused on specific narrative facts of an individual's story, while providing anonymity. Additionally, what was useful about this approach is that the core facts of the narrative, although presented in a different manner, are identical to facts that were found not to qualify for asylum based on the legal standard. This allowed for an objective control for comparison regarding whether the AI and CGC-mediated narrative might elicit more or less empathy than was shown to the actual human asylum seeker. The importance of such research goes beyond the ethics of whether it is just or ethical to portray a narrative with a changed protagonist and medium, as governments are already funding and testing AI algorithms and ML-based CGCs for a variety of purposes, without fully understanding what sort of psychological response CGCs may elicit from humans (see iborderctrl, *Technical Framework*, n.d.; O'Shea et al. 2018; Møhl 2022; Grünenberg et al. 2022). The value of such research transcends artistic research ethics, and involves how experiencers of such media might have potentially long-standing attitudes toward a subaltern class of people shaped toward or away from an "othering" perspective. The outcome of the study seemed to show that AI in the form of ML models such as ChatGPT may indeed offer affordances for reworking narrative forms into more empathetically resonant forms for viewers or experiencers, despite the form of the narrator.

DO SUCH CHANGES, OR TECHNOLOGICAL MEDIATIONS, AFFECT WHETHER A VULNERABLE GROUP WILL BE FAIRLY AND ACCURATELY PORTRAYED?

How one perceives another may be largely trained into a person from hegemonic vantage points that latently imprint into cultural perception a form of relational power. Necessary to explaining empathy, either in the context of real-world immigration settings or in the context of virtual narrative production, requires understanding basic constitutive questions such as what does it mean to be a refugee or to be an "Other"; what does it mean to be a citizen; and what does it mean to be qualified to enter or to belong within a society? AI appears to have the potential to subvert assumptions about identities by reconfiguring narrative facts into translational forms that may possess the ability to alter preconceptions.

Technologically mediated narratives appear to have the capacity to some extent to deconstruct "hegemonic relations" that form the basis for understanding relations of power as constituting apparent social positions and properties (cf. Selg 2013: 9). The law on refugees is a *universal* hegemonic expression of the judgment of legitimacy of people to freely travel, and to belong to or to enjoy rights within a given society. Such a relationship acts to universalize or objectify a popular social perception of identity for peoples through a hegemonic lens of definition; however, such claimed "'social objectivity is constituted through acts of power, … is ultimately political, [and] 'has to show the traces of exclusion which governs its constitution'" (Mouffe 2000: 99, as quoted in Selg 2013: 10). In fact, from a relational perspective, power itself "should not be conceived as an external relation taking place between two preconstituted identities, but rather as constituting the identities themselves" (Mouffe 2000: 99, as quoted in Selg 2013: 10; see also Emirbayer 1997: 291–292). This is evident in the sense that a person called a refugee is only a refugee in the context of an imbalanced power dynamic

that invents and defines the term refugee itself. Refugees are "constitutive of" (Selg, Ventsel 2020: 221) the immigration setting that defines them. The system itself creates "a new status" (ibid.) for those displaced peoples searching for a better life, "making some X [suffering people] count as a Y [a refugee]" (ibid.). Of course, refugees are not the only hegemonically constituted position within the immigration setting. This constitutive relationship can be clearly seen in the case of refugees/asylum seekers and citizens/immigration officers, each defined in terms of power relations with "invidious categorical distinctions" (Emirbayer 1997: 292), mutually arising to each other and the system itself. From this perspective, technological mediation may be seen as subverting such categories, and thereby posing a challenge to existing power structures.

What is critical in this constitutive understanding of the immigration setting and the refugee's place in it is the point that what has been socially constructed may also be socially deconstructed, or undone (Diaz-Leon 2015: 9). If "instantiating a certain category" (ibid.), such as being a *refugee*, is constitutive of "standing in a certain relation to certain social practices [or belief systems], [it is thought then that this fact] opens a clear path for social change: just change those social practices [or beliefs], and social change will automatically follow" (ibid.). An interesting example of such a phenomenon is given by Alexander Wendt (1998: 109) in describing the constitutive relationship of social phenomena to reification through the prism of the Cold War:

> If shared ideas do not explain the Cold War, then policymakers could not end the Cold War by changing their ideas. *The Cold War becomes seen as something external to how they think*, in the same way that rocks and trees are external, and as such can only be dealt with in the manner in which we deal with nature, i.e. through the manipulation and control of objects for whose dispositions *we are not ourselves responsible*. Such *reification* in fact characterized the Cold War for many years; each side thought the conflict was caused by the intrinsically aggressive, implacably hostile nature of the Other, *not by the behaviour of the Self*. *Their own ideas about the conflict were therefore nothing more than necessary reflections of an externally existing fact, not constitutive of that fact in the first place.* (ibid., emphasis added)

At the core of this explanation is the claim that belief systems constitute states of affairs and properties in the world. In this way, interventions capable of causing changes in belief systems – such as altering the way one sees a refugee based on mediating narrative expression through AI – may also change the constitutive status quo, including causal capacities, such as a war's ability to take the lives of people. In this way, "'beliefs and practices are constitutive of each other'" (Rhodes 2017: 18, as quoted in Selg, Ventsel 2020: 224).

In this sense, altering perception of constitutive hegemonic practices and belief systems may allow one to see perhaps the most critical relational aspect, the universality of shared humanity. This reframing or "de-reifying" (Selg, Ventsel 2020: 28) of hegemonically constituted beliefs and perspectives, in the case of people who "count as" refugees (ibid., 220–221) – so that they instead "count as" something more akin to or similar to oneself – might be termed compassion; here empathy may constitute a new relational instance of equivalence, where *refugee* is no longer reified as a "natural kind" but rather seen for what it is, an "ongoing social construction[]" (ibid., 224; see also Wendt 1998: 108–109). In this way empathy itself should be understood not in terms of a trait but rather as a reconceptualization of constitutive relations, practices, and beliefs, reflective of the danger of "'naturalization' of social kinds" as if they were some sort of fixed ontologies (Wendt

1998: 108–109, as quoted in Selg, Ventsel 2020: 224).

AI's capacity to transform narrative to the extent of altering how identity is understood is heavily understudied. Accordingly, the focus of this constitutive explanation is to beg the question whether, and in what way, AI-mediated narratives might causally influence such reconceptualization or reconstitution in their viewers. If certain identities of vulnerable populations are already being unfairly and inaccurately portrayed from a hegemonically defined standpoint, then technologically or artistically mediating such portrayals, with the intention of creating reflexivity toward perception itself, should not be seen as an unethical undertaking.

CONCLUSION: WHAT ARE THE IMPLICATIONS, EITHER WAY?

Despite its promise in altering narratives and thus altering perceptions, which then color the world, AI in particular has spurred many recent flashing warnings and doomsaying caveats. For instance, recently the computer scientist renowned as "the godfather of AI," Dr. Geoffrey Hinton (who created the underlying neural network research foundational to AI models such as ChatGPT), resigned from Google and issued a stark warning: as a result of the technology, not only might "the average person ... 'not be able to know what is true anymore,'" (in the advent of algorithmic creations such as deepfakes), but AI "could actually get smarter than people" and escape human control (Metz 2023). Such predictions perhaps lose focus on the possibility that this may have already occurred, depending on how one defines truth and intelligence. While this discussion is beyond the scope of this paper, it is certainly an area of research that is crucial, and deserves more cross-disciplinary attention, especially in academia.

Matter-of-factly, Hinton, as the creator of the technology, makes it clear that he believes there should be a moratorium on AI growth until societies have "'understood whether they can control it'" (ibid.). Whether it is possible to control something that we do not understand gives weight to the importance of study in the field. Following suit, the European parliament recently passed the Artificial Intelligence Act, which regulates and prohibits certain types of AI activity as "classified under four levels of risk" (Sharp 2023). Of particular note, "AI focused on children ... [and] other vulnerable populations ... would face tougher scrutiny" (ibid.). Such classifications can perhaps only be seen as signposts in the darkness of uncertainty at this point, as it is perhaps one of the most unique situations with which humanity has been confronted at this point – a form of being so profound we term it "intelligence," a mirror that can shape us. Placing a moratorium on technological mediation will inhibit attempts to understand what has now already come into being; it is much like protesting ChatGPT as a newfound scriptwriter – it misses the point and is bound to fail. In an ancient sense, one might say that this new magic has already been unleashed upon the world. We do not deny the existence of forces we observe in the universe; we study them, as we study the world, as we simultaneously study the ecology of ourselves.

Without a doubt, the technology in its current state has many risks. As noted in the paper by Suresh and Guttag (2020), for instance, there are several types of ML bias that might negatively affect outcomes. One example, as described above, is historical bias, which may occur "when there is a misalignment between the world and the [AI] model" (McNamara, Tikka 2023: 22, citing Suresh, Guttag 2020). This type of bias involves situations where virtual models incorporate biases from a hegemonic standpoint, lacking representation of historically excluded people or points of view. This type of bias may arise in a situation where there is no representative of a certain group due to circumstances such as alluded to by Gramsci, involving "subaltern classes," where life has not been documented, represented, or archived due to subjugation of or by a certain group, thus not existing in any dataset. This, of course,

closely ties to what Suresh and Guttag (2020) call "representation" and "evaluation" biases, where datasets

> 'do not equally represent the entire space of images that one cares about' (Suresh and Guttag 2020: 2). Such bias may occur when [AI] has been trained on an incomplete or misrepresentative dataset, where only a portion of a given population has been modelled for analysis, or the model has grown stale based on temporal change. (McNamara, Tikka 2023: 22, citing Suresh, Guttag 2020)

This type of bias might also be thought of as the hegemonic bias whereby models fail to take into account diversity or a multitude of cultural variations that exist within one group (Buolamwini, Gebru 2018).

Such bias mirrors human bias. It is perhaps the bias that exists in the Finnish Immigration decision-maker when deciding the future of an applicant for asylum. It is also like an algorithmic bias that might lead to easily misinterpreting a refugee's narrative as not fitting into a model that is already too narrow. This can be seen in a model that assumes that someone can only be discriminated against or persecuted in a certain way or for a certain reason. This might also assume that a certain identity must be a certain way. What research is showing, however, is that although there are human biases incorporated into AI training data, the interpolation processing of ML language algorithms, such as ChatGPT, also shows promise for overcoming such biases by causing reinterpretation of identities and reconfiguring communication, artistic or otherwise. This is why the humanities and the social sciences need to mobilize in the quest to understand this new creature, and our relationship to it.

AI seems clearly capable of communicating with us in ways we do not yet fully understand. Maarten Coëgnarts refers to the "inferential model of communication" (Coëgnarts 2017: 1, citing Grice 1989), which "proposes a model of communication according to which modes of expression[] (for example, languages, films) are not merely conceived as code systems, but foremost as providers of evidence (that is, manifestations) of an underlying thought" (Coëgnarts 2017: 1, citing Wilson, Sperber 2012: 2; see also Mark Johnson 2007, 2013, as cited in Coëgnarts 2017: 5). These underlying thoughts, in terms of AI, may be analogous to humanity's computerized datasets. It is through its processing of the world of cultural artifacts about humanity and human life stored in our vast but relatively new computerized datasets that AI has the profound ability to reconfigure meaning by amalgamating and interpolating data. It allows those who interact with it not only to experience novelty but to reflect. As Coëgnarts (2017: 5) notes, "[f]orm is considered meaningful insofar [as] it prompts the perceiver of the form to infer thought." In this way, AI may not necessarily be a form of intelligence as much as it is a maker of a new type of meaning or meanings. From the standpoint of the "inferential model of communication," we may come full circle back to the sentiment of the oil painter Gregory Block, that "[w]e the viewers are, in the end, the ultimate artists" (Harwell 2022), and AI as a reflection of ourselves allows us to play with our own possibilities across new artistic realms, and perhaps evolve from our old definitions. At the same time, it can perhaps help us to perceive unity across otherness through such reconfigurations.

Coëgnarts references Rudolf Arnheim's (1954, 1969) "comparison of Camille Corot's figurative painting Mother and Child on the Beach and Henry Moore's non-figurative sculpture Two Forms" as a way of comparing relationships of meaning (Coëgnarts 2017: 6, citing Arnheim 1969: 271–273).

It is thought that each of the visual works of art shares an "analogous structural skeleton and inherent patterns of forces" and "[i]n both cases, the themes of 'protection' and 'concern' are embodied in the way one figure (that is, the mother,



A



B

FIGURE 2. (a) Camille Corot's Mother and Child on the Beach (John G. Johnson collection, Cat, 930, Philadelphia Museum of Art) and (b) Henry Moore's Two Forms (Collection, Museum of Modern Art, New York), as cited in Coëgnarts 2017: 6).

the larger of the two units) bends over and reaches to a second figure (that is, the infant, the smaller of the two units, thus 'holding it down, protecting, encompassing, receiving it'" (Coëgnarts 2017: 6). This is perhaps an example of abstraction in visual art that offers a glimpse into technological mediation as well. It is clear from such comparisons that representation may take on many forms, and many more to be discovered. For if intentionality and emotion can be gleaned from a human-formed sculpture, then why not from a human-formed algorithm or AI? It would perhaps be unethical to ignore this.

DISCLAIMER
If not otherwise stated, all images are reproduced under the terms of fair use for academic purposes. The images are not covered by the Creative Commons Attribution 4.0 International Licence.

REFERENCES

Ayers, John W.; Poliak, Adam; Dredze, Mark; Leas, Eric C.; Zhu, Zechariah; Kelley, Jessica B; ... Smith, Davey M. 2023. "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum." – *JAMA Internal Medicine* 183, 6, 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Bodström, Erna. 2020. "Asylum Decisions as Performances: Intertextuality in Internal Credibility Assessment." – *International Journal of Refugee Law* 32, 4, 623–644. https://doi.org/10.1093/ijrl/eeab001

Brooker, Charlie (writer); Pankiw, Ally (director). 2023, June 15. "Joan Is Awful" (Season 6, Episode 1) [TV series episode]. In B. K. Ali, C. Brooker (executive producers), *Black Mirror*. Broke & Bones.

Buolamwini, Joy; Gebru, Timnit. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." – *Proceedings of Machine Learning Research, Conference on Fairness, Accountability, and Transparency* 8, 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html

Coëgnarts, Maarten. 2017. "Cinema and the Embodied Mind: Metaphor and Simulation in Understanding Meaning in Films." – *Palgrave Communications* 3, 17067. https://doi.org/10.1057/palcomms.2017.67

Davis, Mark H. 1983. "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach." – *Journal of Personality and Social Psychology* 44, 1, 113–126. http://dx.doi.org/10.1037/0022-3514.44.1.113

Diaz-Leon, Esa. 2015. "What Is Social Construction?" – *European Journal of Philosophy* 23, 4, 1137–1152. https://doi.org/10.1111/ejop.12033

Di Palma, Michael; Arcangeli, Elisa; Lattanzi, Davide; Gabbiadini, Alessandro; Galluci Marcello; Cuppini, Riccardo; ... Berlinger, Manuela. 2019. "Heart Rate Variability Reveals the Fight between Racially Biased and Politically Correct Behavior." – *Nature Scientific Reports* 9, 11532. https://doi.org/10.1038/s41598-019-47888-w

Elyoseph, Zohar; Hadar-Shoval, Dorit; Asraf, Kfir; Lvovsky, Maya. 2023. "ChatGPT Outperforms Humans in Emotional Awareness Evaluations." – *Frontiers in Psychology* 14, 1199058. https://doi.org/10.3389/fpsyg.2023.1199058

Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." – *American Journal of Sociology* 103, 2, 281–317. https://doi.org/10.1086/231209

Green, Marcus E. 2002. "Gramsci Cannot Speak: Presentations and Interpretations of Gramsci's Concept of the Subaltern." – *Rethinking Marxism* 14, 3, 1–24. http://marcus-e-green.com/PDFs/Green,%20Marcus%20(2002)%20Gramsci%20Cannot%20Speak.pdf (October 9, 2023)

Grimshaw, Anna. 2005. "Eyeing the Field: New Horizons in Visual Anthropology." – Anna Grimshaw, Amanda Ravertz (eds.), *Visualizing Anthropology*. Bristol: Intellect, 17–30.

Grünenberg, Kristina; Møhl, Perle; Olwig, Karen F.; Simonsen, Anja. 2022. "Issue Introduction: IDentities and Identity: Biometric Technologies, Borders and Migration." – *Ethnos* 87, 2, 211–222. https://doi.org/10.1080/00141844.2020.1743336

Harwell, Drew. 2022, September 2. "He Used AI to Win a Fine-Arts Competition. Was It Cheating?" *Washington Post*. https://www.washingtonpost.com/technology/2022/09/02/midjourney-artificial-intelligence-state-fair-colorado/ (October 9, 2023).

Hoffman, Anna L.; Jonas, Anne. 2017. "Recasting Justice for Internet and Online Industry Research Ethics." – Michael Zimmer, Katharina Kinder-Kurlanda (eds.), *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. New York: Peter Lang.

iborderctrl. *Technical Framework*. https://www.iborderctrl.eu/technical-framework (June 6, 2021).

Keung, Nicholas. 2017, January 5. "Canadian Immigration Applications Could Soon Be Assessed by Computers." – *The Star* https://www.thestar.com/news/immigration/canadian-immigration-applications-could-soon-be-assessed-by-computers/article_af154053-4736-5b59-a319-d7d84f8d7841.html (October 9, 2023).

MacDougal, David. 1998. *Transcultural Cinema*. Princeton, NJ: Princeton University Press.

MacDougal, David. 2001. "Review of Making Forest of Bliss: Intention, Circumstance and Chance in Nonfiction Film. A Conversation between Robert Gardner and Ákos Östör." – *Visual Anthropology Review* 17, 1, 68–85. https://doi.org/10.1525/aa.2002.104.2.651

Markham, Annette N.; Buchanan, Elizabeth. 2012. *AoIR Ethics Guidelines 2012 – Ethical Decision Making and Internet Research, Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. http://aoir.org/reports/ethics2.pdf (October 9, 2023).

McNamara, Robert G.; Tikka, Pia. 2023. "Well-Founded Fear of Algorithms or Algorithms of Well-Founded Fear? Hybrid Intelligence in Automated Asylum Seeker Interviews." – *Journal of Refugee Studies,* 36, 2, 238–270. https://doi.org/10.1093/jrs/feac067

Metz, Cade. 2023, May 1. "'The Godfather of A.I.' Leaves Google and Warns of Danger Ahead." – *New York Times*. https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html (October 9, 2023).

Møhl, Perle. 2022. "Biometric Technologies, Data and the Sensory Work of Border Control." – *Ethnos* 87, 2, 241–256. https://doi.org/10.1080/00141844.2019.1696858

Morris, Rosalind C.; Spivak, Gayatri C. 2010. *Can the Subaltern Speak? Reflections on the History of an Idea*. New York: Columbia University Press.

O'Shea, James; Crockett, Keeley; Khan, Wasiq; Kindynis, Philippos; Antoniades, Athos; Boultadakis, Georgios. 2018. "Intelligent Deception Detection through Machine Based Interviewing." – *IEEE International Joint Conference on Artificial Neural Networks (IJCNN)*. https://doi.org/10.1109/IJCNN.2018.8489392

Scott, James C. 2009. "Hills, Valleys, and States: An Introduction to Zomia." – James C. Scott, *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven, CT: Yale University Press, 1–39.

Selg, Peeter. 2013. "The Politics of Theory and the Constitution of Meaning." – *Sociological Theory* 3, 1, 1–23. https://doi.org/10.1177/0735275113479933

Selg, Peeter; Ventsel, Andreas. 2020. *Introducing Relational Political Analysis: Political Semiotics as a Theory and Method.* Cham: Palgrave Macmillan.

Sharp, Alexandra 2023. "EU Lawmakers Pass Landmark AI Regulation Bill, Foreign Policy." https://foreignpolicy.com/2023/06/14/eu-ai-act-european-union-chatgpt-regulations-transparency-privacy/ (October 9, 2023).

Suresh, Harini; Guttag, John V. 2020. "A Framework for Understanding Unintended Consequences of Machine Learning." https://arxiv.org/pdf/1901.10002v3.pdf (October 9, 2023).

Tangermann, Julian. 2017. *Documenting and Establishing Identity in the Migration Process: Challenges and Practices in the German Context*. Nuremberg: Federal Office for Migration and Refugees.

Tiidenberg, Katrin. 2019. "Research Ethics, Vulnerability and Trust on the Internet." – J. Hunsinger, L. Klastrup, M. Allen (eds.), *Second International Handbook of Internet Research*. Cham: Springer.
Tikka, Pia; McNamara, Robert G.; Gerry, Lynda J.; Kosunen, Ilkka; Kaipainen, Mauri .2021. "Designing Enactive Co-Presence: Second-Order Simulation of Empathy for Artificial Humans in Narrative Contexts." – *PRESENCE: Virtual and Augmented Reality* 30, 149–166. https://doi.org/10.1162/pres_a_00378
Timmermans, Stefan; Tavory, Iddo. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." – *Sociological Theory* 30, 3, 167–186. https://doi.org/10.1177/0735275112457914
United Nations High Commissioner for Refugees (UNHCR). 2019. *Handbook on Procedures and Criteria for Determining Refugee Status and Guidelines on International Protection under the 1951 Convention and 1967 Protocol Relating to the Status of Refugees*.
Wendt, Alexander. 1998. "On Constitution and Causation in International Relations." – *Review of International Studies* 24, 5, 101–118. http://www.jstor.org/stable/20097563

ROBERT MCNAMARA is a PhD candidate in Audiovisual Arts and Media Studies and a Junior Researcher at Tallinn University's Baltic Film, Media and Arts School. He has a background in American criminal law and has done public defence work for nearly a decade in New York state. Additionally, he has degrees related to Audiovisual Ethnography and Eastern Classics. His research focuses on ethical, socio-political, anthropological, and legal aspects of employing tools of virtuality across media. As a member of Tallinn University's Enactive Virtuality Lab, he co-authored journal papers with the team related to social issues surrounding the use of artificial intelligence, machine learning, and virtual avatars.