RESEARCH ARTICLE

Conceptual Engineering and Philosophy of Technology: Amelioration or Adaptation?

Jeroen Hopster[1] ⓘ · **Guido Löhr**[2] ⓘ

Abstract

Conceptual Engineering (CE) is thought to be generally aimed at ameliorating deficient concepts. In this paper, we challenge this assumption: we argue that CE is frequently undertaken with the orthogonal aim of *conceptual adaptation*. We develop this thesis with reference to the interplay between technology and concepts. Emerging technologies can exert significant pressure on conceptual systems and spark 'conceptual disruption'. For example, advances in Artificial Intelligence raise the question of whether AIs are agents or mere objects, which can be construed as a CE question regarding the concepts AGENT and OBJECT. We distinguish between three types of conceptual disruption (conceptual gaps, conceptual overlaps, and conceptual misalignments) and argue that when CE occurs to address these disruptions, its primary aim is not to improve concepts, but to retain their functional quality, or to prevent them from degrading. This is the characteristic aim of CE when undertaken in philosophy of technology: to *preserve* the functional role of a concept or conceptual scheme, rather than *improving* how a concept fulfills its respective function.

Keywords Conceptual Engineering · Conceptual adaptation · Amelioration · Disruption · Misalignment · Preservation · Socially Disruptive Technologies

✉ Jeroen Hopster
   j.k.g.hopster@uu.nl

   Guido Löhr
   g.lohr@vu.nl

1   Ethics Institute, Utrecht University, Utrecht, the Netherlands

2   Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

1 Introduction

According to the dominant view in current scholarship on Conceptual Engineering (CE),[1] CE is primarily aimed at conceptual improvement or amelioration.[2] For example, Herman Cappelen (2020, p. 132) outlines CE as follows:

> Conceptual engineering amounts to the following: It is the project of assessing and then ameliorating our concepts. (…) An epistemological conceptual engineer will assess epistemic concepts with the aim of improving them. A conceptual engineer in moral philosophy will aim to assess and improve our moral concepts. A metaphysical ameliorator will try to improve our core metaphysical concepts.

Cappelen is not alone in holding a view along these lines: the assumption that CE aims at conceptual amelioration is widespread (e.g., Haslanger, 2020; Greenough, 2019; for reviews, see Isaac et al., 2022; Koch et al., 2023).[3] Various paradigmatic examples of CE suggest that this assumption is, in fact, correct. For example, Haslanger (2000) argues that we should associate words like 'woman' and 'man' with different concepts to facilitate the fight for gender equality. Manne (2017) has engineered an arguably more appropriate concept for the word 'misogyny'. Scharp (2013) has proposed to engineer a concept for the word 'truth' that avoids certain paradoxes. Each of these examples refers to a CE project aimed at moral, epistemological, or metaphysical gains.

We do not dispute the view that for several core examples in epistemology, metaphysics, and moral philosophy, CE is best understood as being aimed at conceptual amelioration. However, there is an anomalous set of cases that do not fully conform to this picture. These anomalous cases are particularly prevalent when CE takes place in contexts of technological change (e.g., Danaher & Hopster, 2022; Löhr, 2023a; Veluwenkamp et al., 2022; Veluwenkamp & van den Hoven, 2023). When analyzing, from a descriptive angle, the conceptual pressures induced by Socially Disruptive Technologies (Hopster, 2021), we find that CE is not undertaken with the

---

[1] Conceptual Engineering is typically contrasted with conceptual analysis, the intentional de-construction of concepts into necessary and sufficient conditions, or to sets of inferences we are generally said to be entitled to (cf., Koch et al., 2023). By contrast, conceptual engineers *intervene* in our conceptual systems, for instance by revising existing concepts or by introducing new ones, regardless of whether their proposals reflect the dominant uses of concepts in the linguistic community.

[2] What constitute the objects of CE (the things to be improved) is subject to scholarly dissensus: these objects have been taken to be word meanings (Cappelen, 2018); speaker meanings (Pinder, 2021); mentally represented bodies of information (Isaac, 2020; Machery, 2017); dual contents of concepts (Koch, 2021); inferential relations (Haslanger, 2020; Jorem & Löhr, 2022; Thomasson, 2021); or linguistic entitlements (Löhr, 2021).

[3] Another example is the first sentence of a review article by Isaac et al. (2022, p. 1): "Conceptual Engineering is a branch of philosophy concerned with the process of assessing and improving our concepts. It is motivated by the fact that, sometimes, our conceptual schema must be ameliorated to attain certain beneficial consequences, which may be social, theoretical, political, or otherwise."

principal aim of conceptual amelioration.[4] Rather than *improving* concepts relative to a goal, in contexts of technological disruption, conceptual engineers frequently aim at *preserving* conceptual functions to adapt to new circumstances, irrespective of whether such adaptations qualify as ameliorative.[5] In this article, we develop this notion of conceptual adaptation and argue that it provides us with a richer understanding of the aims of CE, in particular as it occurs in response to tech-induced conceptual disruptions.

Our argument proceeds as follows. In Section 2, we outline the notion of conceptual adaptation and contrast it with the notion of conceptual amelioration. In Section 3, we illustrate the process of conceptual adaptation with three case studies of CE in contexts of technology, arguing that these cases are better understood in terms of conceptual adaptation than in terms of amelioration. In Section 4, we clarify the functionalist understanding of concepts that our account presupposes and suggest that the function of concepts can be technologically mediated. We conclude by summarizing our main claims and by proposing conceptual adaptation as an important topic for future inquiry in philosophy of technology.

2 Conceptual Adaptation and Conceptual Amelioration

While Cappelen's framing of conceptual engineering as "the project of assessing and then ameliorating our concepts" is apt for some cases, it also has its problems and limitations. First, this is because CE often involves arguing for the introduction and elimination of concepts, i.e., the improvement of a *conceptual network* rather than the improvement of a single concept. Second, the question of whether it is even possible to improve the same concept (rather than replacing it, thereby eliminating the old one and introducing a new one) is deeply controversial (cf., Koch et al., 2023). Arguably, a more prudent way of characterizing CE is as the intentional change of conceptual schemata or repertoires: networks of concepts that we can expand or reduce (cf., Jorem & Löhr, 2022; Löhr, 2023c; Thomasson, 2021). We can ameliorate this conceptual repertoire by replacing existing concepts with better ones – better relative to key moral, epistemic, or metaphysical goals.

Building on these two lines of criticism, in this article we argue that even if improvement is part of CE, foregrounding this as the project's core aim offers an incomplete account. It obscures the fact that CE often does not occur with the

---

[4] As our case studies (Section 3) bring out, this holds for studies of CE in academic scholarship and in concrete societal settings where stakeholders discuss the conceptual implications of new and emerging technologies – CE 'in the wild', so to say.

[5] Note that we *do not* exclude the possibility that a conceptual adaptation may also count as ameliorative. Consider for example a quote by Floridi (2019): "The world in which we live seems in great need of all the possible help we can give it, and a constructionist philosophy capable of devising the required concepts that will enhance our understanding may definitely lend a hand, if we can manage to develop it." We take Floridi to talk about what we consider conceptual adaptation as well as amelioration here. Sometimes, we adapt to new environments by introducing new concepts that improve our epistemic relation to the world.

principal aim to improve concepts, but rather to retain the functional quality of conceptual networks and repertoires, and to prevent them from degrading. Doing so is imperative in contexts of technological change and disruption, where emerging technologies put pressure on conceptual repertoires. In response, humans engineer their conceptual repertoire by introducing, modifying, and eliminating concepts – and while doing so may involve ameliorative steps, the aim of improvement is neither what triggers CE, nor what best characterizes the engineering project as a whole. In contexts of tech-induced conceptual disruption, the aim of CE is more appropriately framed in terms of *conceptual adaptation*, than conceptual amelioration: conceptual engineers principally seek to preserve a functional conceptual repertoire, rather than improving individual concepts relative to moral, metaphysical, or epistemic goals.

2.1 What is Conceptual Adaptation?

It is commonplace among philosophers and scientists to regard humans as a Promethean species, naturally equipped to use technologies and culturally adapt to a wide range of environments (e.g., Stiegler, 1998; Henrich, 2016). Our joint conceptual repertoires, too, comprise an important reservoir for cultural adaptation. Among many other things, concepts serve to express and reify norms, which structure human social interaction and cooperation. Yet, reification can also be a burden: novel challenges may arise and call upon us to adjust existing social and moral norms or to articulate new norms. In present-day contexts, such challenges are themselves often technological in nature: emerging technologies frequently confront us with disruptions of entrenched normative orientations and call for a reconsideration of the soundness of existing norms, and possibly for their modification (van de Poel et al., 2023). Since these norms are embellished in our language and concepts, resolving normative disruptions might require – and take shape by means of – conceptual modification (Löhr, 2023b).

Following this exposition, *conceptual adaptation* may be regarded as an important component of present-day cultural evolution: we often witness a two-step dynamic of technology-induced social disruption, followed by a process of normative and conceptual adaptation. In this article, we do not seek to develop this claim by committing ourselves to a specific account of cultural evolution. Instead, we aim to point out that conceptual adaptation might also be regarded as a normative aspiration, and that this aspiration is in fact an important goal of CE: conceptual engineers frequently set out to adapt their conceptual repertoires, as doing so is needed to retain a functional conceptual orientation. Prompted by technological changes, CE frequently occurs in a reactive mode: rather than aiming for improvement, conceptual engineers set out to innovate language and adapt conceptual schemes, whose function is jeopardized by technological changes (Löhr, 2023b).[6] Hence, rather than

---

[6] Conceptual adaptation is reactive in the sense that it occurs in response to social and conceptual disruptions. We add, however, that conceptual adaptations may also respond to disruptions that are anticipated to take place in the future. Hence, adaptation may either be backward-looking (reactive), or forward-looking (proactive).

embarking on CE as a project of improvement, conceptual engineers often embark on it as a project of retaining a functional normative orientation, in the face of impending conceptual degradation.

Consider the rise of AI, which prompts the question of whether AIs are persons or mere objects. This can be construed as a conceptual question regarding concepts like AI, AGENT, and OBJECT, and thus be understood as a question of CE (cf. Himmelreich & Köhler, 2022; Löhr, 2023a). Answering this conceptual question can have important social and normative repercussions: whether AIs are classified as persons or objects has implications for the corresponding social and legal norms. Yet it is not a question raised by conceptual activists, who set out to change deficient concepts they find moral and conceptually deficient, nor is it a question raised by fundamental epistemological or metaphysical concerns. Instead, it is a question prompted by technological changes, which put pressure on the adequacy of concepts that are central to providing moral orientation. CE is called for, but conceptual engineers do not initiate the project; they react to conceptual disruptions that, if left unattended, are likely to yield normative regress.

Conceptual disruptions and adaptations need not be technology-driven: warfare, fake news, or environmental hazards might similarly give rise to social and conceptual disruptions, which call for adaptation (Oimann, 2023). Note, however, that technologies do play a significant role in each of these contexts, and the same holds for conceptual disruption more generally: concepts and conceptual schemes frequently have to be revised and adapted in the face of technological pressures. As new and emerging technologies give rise to new entities, new social practices, and new and changing social norms, they are key triggers of conceptual disruption that call for a CE response in turn (cf., Himmelreich & Köhler, 2022; Veluwenkamp et al., 2022). Therefore, the aim of conceptual adaptation is especially salient in the philosophy and ethics of technology.

2.2 Is Conceptual Adaptation Not Just a Kind of Amelioration?

Our thesis that conceptual adaptation is a key aim of CE in contexts of technological disruption is not meant to negate the relevance of conceptual amelioration. Nor do we mean to sketch two CE projects that are mutually exclusive: we think of conceptual adaptation and amelioration as partly overlapping, yet orthogonal aims. We grant that CE often aims at improvement when it comes to frequently discussed concepts such as WOMAN, MISOGYNY, and TRUTH. These concepts arguably can be improved relative to certain epistemic or moral goals. Similarly, conceptual engineers working in technological contexts often do aim for improvement. However, in addition, there are non-ameliorative aspects of CE that are particularly prevalent in contexts of tech-induced conceptual disruption. Successful conceptual adaptation does not necessarily give us 'better concepts' in the ameliorative sense prevalent in the CE debate. Instead, as we will illustrate in the next section, insofar as adaptation yields improvement, it typically does so with regard to conceptual schemes at large, and in the face of changing conceptual needs. That means that the 'improvement' that conceptual adaptation creates does not pertain to a stable conceptual context at

t-1 (as it does in cases of conceptual amelioration), but to resolving conceptual disruptions that play out over time, and in the face of impending regress at t-2.

We grant, then, that conceptual adaptation may be partly ameliorative: adjusting conceptual repertoires in order to retain a functional normative orientation in the face of impending regress, might itself be regarded as a kind of improvement.[7] However, this is not the same kind of improvement as heralded in current CE scholarship. Rather than embarking on CE to achieve moral, metaphysical, and epistemic gains, conceptual adaptation is prompted by the aspiration to prevent our conceptual systems from degrading. For instance, an argument to the effect that the referent of the concept AGENT should extend to AI does not necessarily qualify as an improvement of the concept AGENT. Instead, it can more broadly be understood as an attempt to overcome conceptual uncertainty or conceptual conflict due to a *conceptual overlap* (e.g., that AI can be conceptualized as both an agent and an object). Such a conceptual overlap (Löhr, 2023b) seems undesirable because it yields uncertainty with respect to the associated social and legal norms – an idea we illustrate in the next section. While there is a sense in which resolving this conceptual conflict can qualify as an improvement, as it is conducive to retaining the functional qualities of our conceptual scheme, this only holds if we take into account the changes in the external environment and the different needs we want our conceptual repertoire in the face of such changes – a context that is overlooked in the framing of CE in terms of amelioration.

In sum, the overall picture of CE in the face of technological disruption is not one of improving an isolated concept relative to certain moral or epistemic goals, but one of preventing *conceptual degradation* of a conceptual scheme. The advent of sophisticated general AI, as well as many other technologies, provokes an impending conceptual regress, in the sense that our existing conceptual system no longer provides clear normative guidelines. In the face of this impending regress, the project that conceptual engineers pursue is one of conceptual adaptation: adjusting conceptual frameworks to overcome conceptual gaps, overlaps, and misalignments.

3 Three Case Studies of Conceptual Adaptation

Technology can be a potent source of novelty and an instigator of social and conceptual disruption. It has been argued, for instance, that social media challenge or even require a revision of the concept of friend (Koch, 2016), that neuroimaging techniques eliminate the concept of free will (Swaab, 2014), that digital Twins may alter the concepts of health and disease (Bruynseels et al., 2018), that future sex robots will put pressure on the concept of consent (Frank & Nyholm, 2017), that artificial womb technologies will challenge the concepts of fetus, mother, and parents (Romanis, 2018), that machines disrupt the concept of identity (Babushkina & Votsis, 2021), and that synthetic biology puts pressure on the concept of life (Preston, 2019).

---

[7]　We thank an anonymous reviewer for pressing us on this point.

In this section, we look in detail at three examples of tech-induced conceptual disruption, which are followed by a process of conceptual adaptation. Each of these cases follows a general schema (Fig. 1). First, technology induces certain social changes. Next, these changes generate a disruption to conceptual schemes, by generating either a *conceptual gap*, a *conceptual overlap*, or a *conceptual misalignment*.[8] Finally, this conceptual disruption calls for CE, for instance by the introduction, revision, or elimination of concepts, or by the preservation (Lindauer, 2020) of a concept that is being put under pressure in the context of a disruption.
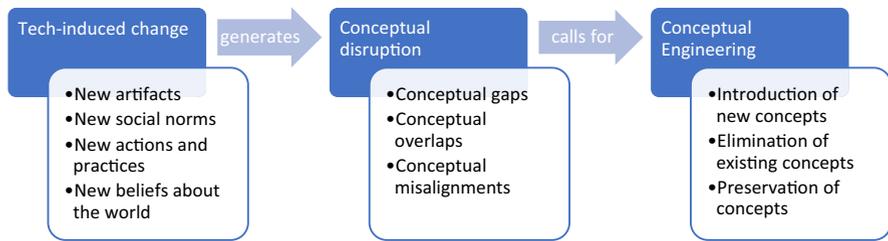
3.1 Mechanical Ventilation and the Concept of Brain-Death

A historical example of CE taking place in society, which has been frequently recounted in recent scholarship on the dynamics between technology and conceptual change (Baker, 2013; Nickel, 2020), concerns the advent of mechanical ventilation technologies and the ensuing disruption of classifications and norms regarding life and death. Mechanical ventilation is the medical use of pumped air to assist patients in breathing (ventilating) when the lungs are not able to pump on their own. This medical technology evolved over the course of the mid-twentieth century and became increasingly effective. By the 1960s, some patients with severe brain injury could retain their lung function indefinitely using mechanical ventilation. The medical state of these patients was without precedent: they showed no significant brain activity but retained the ability to breathe, assisted by ventilation technology. This confronted doctors and families of the patients with classificatory uncertainty: were the patients dead or alive? Descriptive ambiguity about the patients' states was intertwined with moral uncertainty regarding the moral obligations towards them, such as uncertainty as to whether the organs of the patients might be used for organ transplantation, or whether removing ventilation would be an instance of "killing" or "letting die" (Nickel, 2020). What ultimately resolved this moral and conceptual ambiguity was the introduction of a novel concept – BRAIN-DEATH – and the formulation of specific ethical codes associated with this concept, which provided action-guiding rules and norms to both doctors and families.

This case of introducing a new concept is usefully understood as an instance of CE 'in the wild', aiming at conceptual adaptation.[9] The need for such adaptation emerged when novel technology facilitated a new medical state, which lacked a clear descriptor and associated norms – a situation we identify as a *conceptual gap*. The lack of tailored concepts and action-guiding norms to handle this novel state provoked moral uncertainty among relevant bystanders. Arguably this uncertainty was harmful (Nickel, 2020) and can be viewed as an instance of impending

---

[8] We do not claim that this threefold typology of conceptual disruption is exhaustive. Arguably, 'conceptual appropriation' constitutes a further type of conceptual disruption (Hopster et al., 2024), and there may be other types still.

[9] By CE 'in the wild' we mean to denote CE as it takes place in societal settings, which do not involve the intervention of self-identified conceptual engineers.

Tech-induced change		Conceptual disruption		Conceptual Engineering
•New artifacts •New social norms •New actions and practices •New beliefs about the world	generates →	•Conceptual gaps •Conceptual overlaps •Conceptual misalignments	calls for →	•Introduction of new concepts •Elimination of existing concepts •Preservation of concepts

Fig. 1 A schematic overview of the three steps involved in the adaptation of human conceptual schemes, with some of their main types

conceptual regress: due to technological changes, an important epistemic function of the existing conceptual repertoire – namely, to clearly distinguish between patients that are dead or alive (with all its social consequences) – was no longer fulfilled. Therefore, an adaptation of the conceptual scheme was called for. This adaptation transpired when the concept BRAIN-DEATH was introduced and new norms about the treatment of brain-dead patients were articulated, thereby re-establishing a good fit between the affordances of the conceptual repertoire and conceptual needs arising from social practice.

Note that this conceptual adaptation did not involve the *improvement* of any individual concept, such as the concept of DEATH. Arguably, the adaptation did constitute an improvement of the conceptual repertoire as a whole, yet only with respect to the period of conceptual disruption that ensued after the advent of mechanical ventilation technology. Prior to the advent of this technology, there was no need to articulate the concept of BRAIN-DEATH. More precisely put, the introduction of BRAIN-DEATH at t3 constituted an improvement of the medical conceptual system with respect to the moment of conceptual uncertainty at t2, but it did not obviously constitute an improvement concerning the prior state at t1 when mechanical ventilation technologies had not yet been developed. Hence, from t1 to t3, the medical conceptual scheme adapted to a new techno-social reality, and while this process of adaptation involved an element of amelioration from t2 to t3, it also involved a moment of conceptual disruption and maladaptation between t1 and t2. The overall dynamics from t1 to t3 are naturally construed in terms of conceptual adaptation: humans adapted their conceptual repertoire to respond to new social circumstances appropriately and effectively.

3.2 Social Robots and the Concept of Love

Our second case study concerns an example of CE in response to an anticipated future disruption. This time around, CE takes place in the context of a scholarly debate, which involves self-conscious reflection on the content and application of concepts. The concept at issue is (romantic) LOVE and the question is whether this concept could be appropriately ascribed to intimate human–robot relationships. While this may not yet be an issue of major concern in contemporary societies, conceivably it will be in the future, for instance, if robot engineering companies manage

to produce highly sophisticated 'love robots' for the consumer market (as in the movie *Her*). The question of whether intimate human–robot interactions qualify as relations of mutual love could get entangled, either directly or indirectly, with various social and legal norms – for instance, whether humans and love robots can enter into a registered partnership, or whether 'adultery' in human–robot relationships is morally objectionable. Hence, for ethicists, it seems relevant to inquire whether human–robot relationships can give rise to mutual love, in the real sense of LOVE.

To simplify, we might discern two positions in this debate. The first assumes a behaviorist understanding of love (Levy, 2008): If a robot lover speaks and behaves in the same way as a human lover, then we should take this to be an instance of genuine love. The second position is to assume an intention-centric conception of love: What goes on 'inside' robot agents – their emotions, motivations, and intentions – is crucial to ascertain whether or not we can speak of robot love. Endorsing an intention-centric conception, technology ethicists Nyholm and Frank (2017) argue that assuming the present state of robot technology, human–robot relationships do not satisfy the conceptual criteria of love. Furthermore, they suggest that an intention-centric conception of love is preferable in ethical terms, noting that "at least at this current stage, a sex and companionship robot to which a person gets so emotionally attached that the person would hesitate to seek human romantic partners could sensibly be thought to impoverish that person's life by serving as a poor substitute for what would be a fuller and more multidimensional type of relationship." (Nyholm & Frank, 2019, p. 414).

This case study is usefully understood as a case of *conceptual overlap*. The emerging technology of love robots enables practices that can be conceptually classified and interpreted in two or more competing ways. Furthermore, future techno-social change might prompt agents to choose between one of these conceptions and to revise their conceptual scheme accordingly. Plausibly, if left to the robot industry, the behaviorist conception is likely to become more dominant in conceptual schemes. After all, it will benefit this industry if what appears to be loving relations between humans and robots, could be classified as a genuine instance of mutual love. In response, conceptual engineers might follow Nyholm and Frank (2019) and argue that we should adapt our conceptual repertoire in a way that favors a behaviorist understanding of love.[10]

Both positions advance specific kinds of conceptual adaptation, which are associated with further conceptual and social changes. Whether or not the concept of love can be attributed to robots is entangled with the question of whether robots should be regarded as PERSONS, whether LOVE is intrinsically bound to INTENTIONALITY, etc. But note that while both positions argue for conceptual adaptations in the face of (anticipated) conceptual disruptions, the view that romantic LOVE should only be reserved for describing intimate human relationships also involves a

---

[10] Note that this case study differs from the case of mechanical ventilation technologies in that CE occurs *pro-actively*. That is to say, scholarly debate does not address a current conceptual disruption, but responds to an anticipated disruption, which may be fostered by future technological changes and arguments advanced by the love robot industry.

prominent aspect of *conceptual preservation* (Lindauer, 2020). Ethicists advancing this view are likely to argue that the currently dominant concept of love is intention-centric and that there are good reasons to keep it that way.

3.3 Large Language Models and the Concept of Agency

The first two case studies call for conceptual adaptation to resolve a conceptual gap or a conceptual overlap.[11] Tech-induced *conceptual gaps* emerge when we lack the conceptual resources to classify and conceptualize novel technologies, yielding conceptual uncertainty. Tech-induced *conceptual overlaps* emerge when more than one concept fits the novel technology (or associated products, norms, etc.). In some cases, different concepts or conceptions can co-exist; different renderings of a concept might be fit for different purposes in different domains. However, conceptual disagreements can also be a source of conceptual uncertainty and may give rise to conceptual disputes, which call for a resolution in the form of a conceptual adaptation.[12]

In addition to conceptual gaps and conceptual overlaps, a third trigger of conceptual adaptations are *conceptual misalignments* (see Fig. 2 for an overview). By this, we mean situations in which a given concept that is entrenched in a joint conceptual scheme is insufficiently aligned with the overall goals (prudential, moral, etc.) of the agents who deploy it. Like instances of a conceptual gap, there is a mismatch between how agents want their conceptual scheme to function and the expressive power that the conceptual scheme allows for in practice. But contrary to the earlier example of a conceptual gap, the problem is not a complete *absence* of the needed concept, but rather the *presence* of a related concept that is supposed to fulfill an important function but fails to adequately do so. For instance, the concept may be widely used, but it does not provide the kind of normative guidance that would be desirable. As a result, conceptual adaptation is called for: the conceptual scheme should be amended to provide the guidance that is needed.

Consider the concept of agency and appeal to it in the recent discourse of generative AI, and more specifically Large Language Models (LLMs). In Spring 2023, the Future of Life Institute published an open letter calling for a "pause in the training of AI systems more powerful than GPT-4" (Future of Life Institute, 2023). Underlying their proposed R&D ban was the concern that "AI labs [are] locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control", which might have "potentially catastrophic effects on society." Some critics, in turn, argued that the institute's worries about unaligned AI were overblown. For instance, a prominent group of AI ethicists argued that "[i]t is dangerous to distract ourselves with a fantasized AI-enabled utopia or apocalypse, which promises either a "flourishing" or "potentially

---

[11] We build on Crootof and Ard's (2021) terminology of 'gaps' and 'overlaps' in the legal domain (we added the third category of misalignment). Our threefold distinction has been taken up by Van de Poel et al. (2023).

[12] Conceptual adaptation takes place in various contexts of discourse, including ethical and legal contexts. See Hopster and Maas (2023).

Concept	Domain	Challenge	Resolution	Normative aim
BRAIN-DEATH	Medicine	Resolving a conceptual gap	Introducing a novel concept	Protecting patients' rights
LOVE	Human-robot interaction	Resolving a conceptual overlap	Preserving an intention-centric conception	Facilitating human wellbeing
AGENCY	Corporate responsibility	Resolving a conceptual misalignment	Preserving a human-centric conception	Holding AI-developers accountable

Fig. 2 Schematization of three case studies of conceptual disruption, followed by conceptual adaptation

catastrophic" future. Such language that inflates the capabilities of automated systems and anthropomorphizes them (…) deceives people into thinking that there is a sentient being behind the synthetic media." (DAIR Institute, 2023).

Importantly, the critics emphasized that the language used by the Future of Life Institute also involved a misleading conceptual claim: they argued that the open letter "not only lures people into uncritically trusting the outputs of systems like ChatGPT but also misattributes agency. Accountability properly lies not with the artifacts but with their builders." This statement could be reframed as an allegation of *conceptual misalignment*: the Future of Life Institute misconstrues the concept of agency, which is suggestive of various other capacities and responsibilities that cannot be properly ascribed to LLMs. Such misalignment is morally consequential, as it can be part of a strategy of AI developers to divert responsibility. The critics argue that this strategy is morally dubious: ascribing the concept AGENCY to LLMs leads to a situation where our conceptual scheme is insufficiently aligned with our overall normative aims – more specifically, to hold AI developers accountable, rather than diverting responsibility to the AI systems themselves.[13]

4 A Functionalist Approach to Conceptual Adaptation

What is the point of adapting our conceptual system to new technology-induced changes or otherwise changed circumstances? In the foregoing sections, we have gestured at the 'social role' or 'function' that we want concepts to serve, which may require adjustment in response to external changes. Let us make this 'function-talk' more precise.

---

[13] Another claim that has been made about the conceptual implications of LLM's is that they decouple agency from intelligence (Floridi, 2023).

A central assumption of the account we propose is that concepts frequently serve as tools that help their users realize normative goals.[14] Some of these goals are social: concepts may serve to signal and perpetuate social norms of all sorts, and conceptual adaptation may serve as a corrective to existing norms. Normativity is deeply entrenched in our conceptual system, often in ways that may not be immediately apparent. For instance, to refer to a piece of cloth as a 'curtain' implicitly alludes to a set of norms about how this piece of cloth is appropriately used; to refer to a person as a 'friend' implicitly suggests social norms about how this person should be treated. Our claim that concepts have a social function should be understood accordingly: it is meant to indicate that concepts serve to signal and perpetuate norms of all sorts and are essential in coordinating joint action (cf. Gibbard, 1990). Conversely, disputes over the norms that should be followed in a given context of joint action can play out at the level of concepts. The same holds for social and institutional interactions more generally: these are regulated by norms, which are entrenched by certain key concepts. Normative disputes may occur as disputes over whether or not the concept of privacy applies to a given practice of data collection,[15] or about how the concept of agency is tied to the concept RESPONSIBILITY. Given that concepts are associated with sets of social norms and practices, arguments for conceptual change go hand in hand with arguments for social change.

Concepts do not only serve as *social* but also as *epistemic* tools (Cappelen, 2018; Isaac, 2021; Simion, 2018). These roles can be related. One epistemic function of concepts is to assist us in thinking and talking about objects or relations in the world. If a conceptual scheme no longer fulfills this epistemic function in some relevant respect, then it should be amended. Like sharpening a knife that has lost its cutting function, so we can 'sharpen' our concepts to make them function in desired ways. Consider our first case study, which described how the distinction between life and death was challenged when the mechanical ventilator was introduced. This loss of epistemic clarity came along with a loss of moral clarity, as the concepts of life and death are associated with a specific set of moral norms. To overcome this ambiguity, the new complex concept BRAIN DEATH was introduced. This introduction also served to preserve the clarity of the distinction between life and death: both concepts could now be continued to be applied as before, as the tech-induced ambiguity was resolved.

The point of departure of CE aiming at conceptual adaptation is to think of the function of our conceptual repertoire as a whole, in terms of a network- or systems analysis. In such an analysis, concepts are regarded as part of a broader network or system. This system may be disrupted, which happens, for instance, when important concepts that are central to the system's epistemic and action-guiding potential, such as AGENCY, LOVE, and DEATH, are challenged and the norms associated with these concepts become muddled. In response to such disruptions, we need to

---

[14] Functions yield normative standards – while these may be *moral* standards, they can also yield normative standards that are independent from morality (Burelli, 2022).

[15] PRIVACY is a likely candidate for frequent conceptual contestation, as its meaning appears to be essentially contested (Mulligan et al., 2016).

amend these and related concepts to bring the system back into a state of equilibrium and restore its normative role. An illustration of this is the *conceptual overlap* we described in our case study on the concept of love. Imagine we broaden the concept of love such that it can be applied to human–robot interactions. This conceptual change puts pressure on related concepts: for instance, it implies that we either have to broaden the concept INTENTION such that we can apply an intentions-centered notion of love, or it means that we have to change the concept LOVE such that intentions are no longer required for it. Conceptual adaptation can thus be thought of as bringing a conceptual system back into equilibrium by resolving conceptual tensions, or by establishing a conceptual system that serves some desired function.

To think about concepts as serving a functional role within a larger system of interrelated concepts that may be destabilized also elucidates the earlier example of a *conceptual misalignment*. The concept AGENCY is closely tied to the concept RESPONSIBILITY: some kind of autonomous agency is typically presupposed for ascriptions of responsibility. Yet agency may not suffice for responsibility: arguably, some human-like capacity for reason responsiveness, not present in current generation AI systems, is additionally required for holding an agent responsible (see van de Poel & Sand, 2018 on conditions for ascribing responsibility). If agency is ascribed to AI systems, whereas such systems lack the requisite kind of reason responsiveness, this may result in a responsibility gap: a situation where no one can be held responsible for the harms of an autonomous system (Matthias, 2004; Oimann, 2023). Such a situation seems undesirable: we want to possess a conceptual and normative system that allows for justified ascriptions of responsibility, especially where powerful Socially Disruptive Technologies (SDTs) are at play. Accordingly, preserving a functional conceptual repertoire arguably involves preserving a repertoire that does not allow for responsibility gaps in the face of SDTs. If such responsibility gaps emerge through a conceptual misalignment, or if the risk thereof becomes manifest, then conceptual adaptation is called for, as suggested by the example of LLMs and the concept AGENCY.

We take the account of conceptual function we have outlined here to be largely uncontroversial. We are aware of more specific proposals about the nature of conceptual function that have been advanced in the CE literature (e.g., Nado, 2021; Thomasson, 2021). Yet, our observations on the social and epistemic function of concepts do not require us to commit to any specific position in this debate. Furthermore, like Riggs (2021) and Jorem (2022), we contend that the notion of conceptual function is probably best understood in deflationary terms: function talk gestures at "what it is about a concept that matters to us in a particular context." (Riggs, 2021).

In closing this section, we highlight one insight implicit in the preceding discussion: the social function that concepts serve is technologically mediated. What social functions concepts serve, and which functions we want them to serve, are interwoven with the technological constellation in which society finds itself. This point seems underappreciated in scholarship on philosophy and ethics of technology: while mediation theory and moral mediation (Verbeek, 2011) are popular and much-discussed theoretical frameworks in recent technology ethics, existing expositions of this framework do not highlight that technologies do not only mediate social and moral life but also mediate our concepts (cf. Coeckelbergh, 2017). Arguably,

conceptual mediation is a promising line of inquiry for future studies of conceptual adaptation, as the technological mediation of concepts is intricately tied to processes of conceptual stability and change. We noted that conceptual stability emerges by virtue of functional stability: if a conceptual repertoire fulfills the functions that agents want it to fulfill, then conceptual schemes will tend to stabilize. If, on the other hand, there is a misalignment between concepts' desired function and their actual function, then conceptual adaptation is called for and change may ensue. As we have seen, such misalignment is frequently due to technological changes in society. Technological disruptions may alter the function of our current conceptual repertoire or may generate new socio-conceptual needs, which the existing conceptual repertoire fails to satisfy. Hence, technology plays into the social functions that we want our concepts to serve. Clarifying the exact nature and varieties of this conceptual mediation may help us to get a better grasp on the processes of conceptual stability and change.

5 Conclusion

CE often occurs as a response to conceptual pressures and disruptions provoked by new and emerging technologies, in a deliberate attempt to adapt conceptual frameworks to new socio-technical environments. We have argued that the project of conceptual adaptation does not coincide with the project of conceptual amelioration. Highlighting conceptual adaptation as a distinct project of CE illuminates relevant phenomena that are largely overlooked by focusing on amelioration. One such phenomenon is conceptual preservation: conceptual engineers may push back against tech-induced conceptual changes that trigger conceptual instability, by seeking to preserve important concepts. That is, conceptual engineers may respond to (anticipated) disruptions by re-asserting concepts, thereby stabilizing or preserving conceptual schemes, in order to prevent them from degrading. Conceptual engineers – either self-proclaimed CE scholars or actors 'in the wild' – adapt conceptual schemes to overcome conceptual gaps, resolve conceptual overlaps, or address conceptual misalignments. Conceptual adaptation is a way of reasserting or altering social norms and expectations, which is often called for in the wake of technological disruption.

We have not argued *against* conceptual amelioration as an important project of CE. Instead, we have proposed that CE has a complementary outlook, which is particularly prominent in the philosophy of technology. Thinking of tech-induced conceptual change in terms of adaptation poses theoretical as well as practical advantages for conceptual engineers. It shifts focus to the complex and entangled dynamics of conceptual change and underscores the reactive mode in which CE typically takes place in the face of SDTs. Furthermore, it facilitates a more tailored ethical palette to intervene in these dynamics (cf. Brun, 2022), including the conservative project of maintaining concepts that are challenged by technological disruptions. Much recent scholarship in technology ethics already engages in conceptual adaptation, but often not self-consciously so. We believe that the analysis we have provided and the conceptual tools we have articulated, such as 'conceptual gap', 'conceptual

overlap', 'conceptual misalignment', and 'conceptual preservation', are beneficial for future undertakings in this field. Working out this theory of conceptual adaptation in more detail, specifically by elucidating how technology mediates our conceptual schemes, is a promising topic for future inquiry in technology ethics.

Authors' Contributions This paper has been a collaborative effort, which emerged from joint discussions. Both authors took part in discussing the paper's contents and structure. J.H. conceived of the notion of conceptual adaptation; G.L. contrasted it with conceptual amelioration. Both authors have contributed to the writing and editing of the final manuscript.

Data Availability Not applicable. No data had to be collected for this research.

Declarations

Ethics Approval and Consent to Participate Not applicable. No ethics approval had to be requested for this research.

Consent for Publication Both authors consent to publication.

Competing Interests Authors declare they have no competing interests.

References

Babushkina, D., & Votsis, A. (2021). Disruption, technology and the question of (Artificial) identity. *AI and Ethics, 2*(4), 611–622. https://doi.org/10.1007/s43681-021-00110-y

Baker, R. (2013). *Before bioethics*. Oxford University Press.

Brun, G. (2022). Re-engineering contested concepts. A reflective-equilibrium approach. *Synthese, 200*(2), 168.

Bruynseels, K., Santoni de Sio, F., & van den Hoven, J. (2018). Digital Twins in Health Care: ethical implications of an emerging engineering paradigm. *Frontiers in Genetics, 9*, 31. https://doi.org/10.3389/fgene.2018.00031

Burelli, C. (2022). Political normativity and the functional autonomy of politics. *European Journal of Political Theory, 21*(4), 627–649.

Cappelen, H. (2018). *Fixing language: An essay on Conceptual Engineering*. Oxford University Press.

Cappelen, H. (2020). Conceptual engineering: the master argument. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 132–151). Oxford University Press.

Coeckelbergh, M. (2017). Language and technology: Maps, bridges, and pathways. *AI & Society, 32*, 175–189.

Crootof, R., & Ard, B. J. (2021). Structuring techlaw. *Harvard Journal of Law & Technology, 34*(2), 347–417. https://doi.org/10.2139/ssrn.3664124

DAIR Institute. (2023). Statement from the listed authors of Stochastic Parrots on the "AI Pause" Letter. https://www.dair-institute.org/blog/letter-statement-March2023. Accessed 20 Apr 2023

Danaher, J., & Hopster, J. K. G. (2022). The normative significance of moral revolutions. *Futures, 103046*, 1–15. https://doi.org/10.1016/j.futures.2022.103046

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(1), 15.

Frank, L., & Nyholm, S. (2017). Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable? *Artificial Intelligence and Law, 25*(3), 305–323.

Future of Life Institute. (2023). Pause Giant AI Experiments: An Open Letter (March 28th 2023). https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.

Greenough, P. (2019). Conceptual marxism and truth: *Inquiry* symposium on Kevin Scharp's *Replacing Truth*. *Inquiry, 62*(4), 403–421. https://doi.org/10.1080/0020174X.2017.1287919

Haslanger, S. (2000). Gender and Race: (What) are they? (What) do we want them to be? *Noûs, 34*(1), 31–55.

Haslanger, S. (2020). Going on, not in the same way. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 230–260)

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Himmelreich, J., & Köhler, S. (2022). Responsible AI through conceptual engineering. *Philosophy & Technology, 35*(60), 1–30. https://doi.org/10.1007/s13347-022-00542-2

Hopster, J. K. G. (2021). What are socially disruptive technologies? *Technology in Society, 67*(101750), 1–8. https://doi.org/10.1016/j.techsoc.2021.101750

Hopster, J. K. G., & Maas, M. (2023). The technology triad: Disruptive AI, regulatory gaps and value change. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00305-5

Hopster, J. K. G., Gerola, A., Hofbauer, B., Löhr, G., Rijssenbeek, J., & Korenhof, P. (2024). Who owns 'nature'? Conceptual Appropration in discourses on Climate- and Biotechnologies. *Environmental Values (forthcoming)*.

Isaac, M. G. (2020). How to conceptually engineer conceptual engineering. *Inquiry, 1*–24. https://doi.org/10.1080/0020174x.2020.1719881

Isaac, M. G. (2021). Post-truth conceptual engineering. *Inquiry, 1*–16. https://doi.org/10.1080/0020174X.2021.1887758

Isaac, M. G., Koch, S., & Nefdt, R. (2022). Conceptual engineering: A road map to practice. *Philosophy Compass*, e12879. https://doi.org/10.1111/phc3.12879

Jorem, S. (2022). The good, the bad and the insignificant – assessing concept functions for conceptual engineering. *Synthese, 200*(2), 1–20.

Jorem, S., & Löhr, G. (2022). Inferentialist conceptual engineering. *Inquiry, 1*–22. https://doi.org/10.1080/0020174X.2022.2062045

Koch, P. (2016). Meaning change and semantic shifts. In P. Juvonen & M. Koptjevskaja-Tamm (Eds.), *The lexical typology of semantic shifts* (pp. 21–66). De Gruyter Mouton. https://doi.org/10.1515/9783110377675-002

Koch, S. (2021). Engineering what? On concepts in conceptual engineering. *Synthese, 199*(1), 1955–1975.

Koch, S., Löhr, G., & Pinder, M. (2023). Recent work in the theory of conceptual engineering. *Analysis*. https://doi.org/10.1093/analys/anad032

Levy, D. (2008). *Love and sex with robots*. Harper.

Lindauer, M. (2020). Conceptual engineering as concept preservation. *Ratio, 33*(3), 155–162.

Löhr, G. (2021). Commitment engineering: Conceptual engineering without representations. *Synthese, 199*(5), 13035–13052.

Löhr, G. (2023a). If conceptual engineering is a new method in the ethics of AI, what method is it exactly? *AI & Ethics, 1*–11. https://doi.org/10.1007/s43681-023-00295-4

Löhr, G. (2023b). Conceptual disruption and 21st century technologies: A framework. *Technology in Society*. https://doi.org/10.1016/j.techsoc.2023.102327

Löhr, G. (2023c). Do socially disruptive technologies really change our concepts or just our conceptions? *Technology in Society, 72*, 102160. https://doi.org/10.1016/j.techsoc.2022.102160

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.

Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*, 175–183.

Mulligan, D. K., Koopman, C., & Doty, N. (2016). Privacy is an essentially contested concept: A multi-dimensional analytic for mapping privacy. *Philosophical Transactions of the Royal Society A, 374*(2083), 20160118.

Nado, J. (2021). Conceptual engineering, truth, and efficacy. *Synthese, 198*, 1507–1527.

Nickel, P. J. (2020). Disruptive Innovation and Moral Uncertainty. *NanoEthics, 14*(3), 259–269.

Nyholm, S., & Frank, L. E. (2017). From sex robots to love robots: Is mutual love with a robot possible? In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications* (pp. 219–45). The MIT Press. https://doi.org/10.7551/mitpress/9780262036689.003.0012

Nyholm, S., & Frank, L. (2019). It loves me, it loves me not: Is it morally problematic to design sex robots that appear to love their owners? *Techné: Research in Philosophy and Technology, 23*(3), 402–424.

Oimann, A. K. (2023). The responsibility gap and LAWS: A critical mapping of the debate. *Philosophy & Technology, 36*(1), 3.

Pinder, M. (2021). Conceptual engineering, metasemantic externalism and speaker-meaning. *Mind, 130*(517), 141–163.

Poel, I. van de, & Sand, M. (2018). Varieties of responsibility: Two problems of responsible innovation. Synthese, pp. 1–20. https://doi.org/10.1007/s11229-018-01951-7.

Preston, C. J. (2019). *The synthetic age: Outdesigning evolution, resurrecting species, and reengineering our world*. MIT Press.

Riggs, J. (2021). Deflating the functional turn in conceptual engineering. *Synthese, 199*, 11555–11586. https://doi.org/10.1007/s11229-021-03302-5

Romanis, E. C. (2018). Artificial womb technology and the frontiers of human reproduction: Conceptual differences and potential implications. *journal of Medical Ethics, 44*(11), 751–755.

Scharp, K. (2013). *Replacing truth*. Oxford University Press.

Simion, M. (2018). The 'Should' in conceptual engineering. *Inquiry, 61*(8), 914–928.

Stiegler, B. (1998). *Technics and time, 1: The fault of Epimetheus* (Vol. 1). Stanford University Press.

Swaab, D. F. (2014). *We are our brains: A neurobiography of the brain, from the womb to Alzheimer's*. Random House.

Thomasson, A. (2021). Conceptual engineering: When do we need it? How can we do it? *Inquiry,* 1–26. https://doi.org/10.1080/0020174X.2021.2000118

van de Poel, I., et al. (2023). *Ethics of socially disruptive technologies: An introduction*. Open Book Publishers. https://doi.org/10.11647/OBP.0366

Veluwenkamp, H., & van den Hoven, J. (2023). Design for values and conceptual engineering. *Ethics and Information Technology, 25*(1), 1–12.

Veluwenkamp, H., Capassa, M., Maas, J., & Lavin, M. (2022). Technology as driver for morally motivated conceptual engineering. *Philosophy & Technology, 35*, 71.

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.