Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A situated conceptual methodology for AI-induced conceptual disruptions

Céline Henne

c.l.l.henne@vu.nl

Vrije Universiteit Amsterdam

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11569852/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A situated conceptual methodology for AI-induced conceptual disruptions

ABSTRACT

Recent advances in AI chatbots and robotics have generated uncertainty and disagreement about the application of many familiar concepts such as responsibility, authorship, creativity, or friendship. Which philosophical methods are most fruitful in the face of such conceptual disruptions? Two approaches currently dominate the current literature. Conceptual analysis seeks to identify the content of the relevant concept and uses it to determine how that concept applies to novel phenomena. Conceptual engineering instead asks which existing or newly constructed concepts would be best suited to the new conditions. In this paper, I develop a distinctive argument against the use of conceptual analysis to resolve conceptual disruptions, even for those who do not believe in the automatic superiority of conceptual engineering over conceptual analysis. The challenge rests on the idea of open texture: nothing in the content of concepts determines in advance how they should be applied in every novel situation; instead, because concepts only work effectively within stable forms of life, any disruption in those background regularities has the potential of making the conceptual practice incoherent or unstable, in a way that cannot be resolved by looking at how things were done in pre-disrupted times. I then defend a positive proposal for addressing conceptual disruptions, in several steps: situated conceptual analysis diagnoses disruptions by examining concepts alongside the forms of life that enable their effective operation; situated conceptual ethics and engineering imagine and evaluate possible continuations of conceptual practices together with the forms of life that would sustain them.

Whether they are caused by the discovery of new phenomena or the advancement of new technologies, conceptual disruptions arise when our conceptual practices become unsettled in the face of novelty (Löhr 2023a). Recently, the advent of sophisticated LLM chatbots and robotics has generated uncertainty and disagreement with regard to the way we apply concepts that were, on the whole, unproblematically applied to human-only phenomena. We are unsure who is *responsible* for the actions of an autonomous AI system, whether someone can have an AI *friend*, whether we can *trust* or only *rely on* an LLM for decision-making, or whether ChatGPT can be an *author*. Some of these conceptual questions are pressing, because practical consequences hang on their resolution. For example, the resolution of the question regarding responsibility will determine who may be blamed for the actions of these systems, whether and how compensation can be sought, and will influence the institutional and engineering practices surrounding the design and control of these systems.

Philosophy is the discipline that deals most explicitly with conceptual questions. This paper asks the following question: What philosophical methods are most helpful in understanding and addressing these conceptual disruptions?

Two competing methods are currently prominent in the philosophy of AI. Conceptual analysis, which has long served as analytic philosophy's default method for addressing conceptual questions, aims at revealing or clarifying the content of concepts or the meanings of words. Unsurprisingly, it has become a method of choice for dealing with the kinds of conceptual puzzles caused by the rapid advancement and sophistication of digital technologies. If we want to know whether relationships with AI systems could count as friendships, we should first establish what "friendship" means, or what counts as friendship in general, and then examine whether human-AI relationships could satisfy these conditions (Danaher 2019; Fröding and Peterson 2021; Munn and Weijers 2025; Lott and Hasselberger 2025). This is what I call the *descriptive-analytic methodology*, in which the product of conceptual analysis occupies the central argumentative role in resolving conceptual questions.

Many philosophers, however, regard conceptual engineering as the more appropriate method for addressing such questions (Himmelreich and Köhler 2022; Hopster and Löhr 2023; Löhr

2023b; Viehoff 2023; Veluwenkamp et al. 2024). While the descriptive-analytic methodology treats the content of the concept we currently possess as providing the authoritative basis for classifying the novel case, what I call the *normative-engineering methodology* asks which concept or classification we have reason to adopt or best serves our purposes, even if this implies a departure from the original concept. Existing defences of conceptual engineering as the method of choice in philosophy of AI (or in other domains) commonly emphasize the normative gap between the concepts we currently possess and the concepts we ought to use. Since our inherited concepts are not necessarily the best concepts we could have, and are the kinds of things that can be evaluated, changed and improved, these authors argue that a normative rather than descriptive method is better suited to address conceptual questions.

In this paper, I do two things. I develop a different and more specific challenge to the descriptive-analytic methodology, which is specifically targeted at technology-induced conceptual disruptions. Then, I draw the consequences of this challenge and defend a methodological proposal that, I will argue, is best adapted to deal with conceptual disruptions.

The challenge to the descriptive-analytic methodology is the following. I argue that most conceptual disruptions are due to the *open texture* of concepts. Open texture is the idea that nothing in the content of concepts determines in advance how they should be applied in all novel situations. Our conceptual practices were formed under conditions in which no AI systems existed, and cannot supply definite rules for how to conceptualize or classify these novel phenomena. Accordingly, analysis of our current concepts cannot, by itself, be used to resolve these conceptual questions. Unlike the standard normative-gap argument, this challenge does not presuppose that conceptual engineering is generally superior to conceptual analysis.

Rather than immediately turning to the normative-engineering methodology as it is practised, I draw another important methodological lesson from open texture and its manifestation in conceptual disruptions. What these events reveal is not only that concepts are incomplete, but also that they do not live in vacuums: their effective operation depends on *forms of life*, understood as the webs of material and institutional circumstances, capacities, and concerns

that make our concepts intelligible and effective. As these conditions change, criteria of application that once converged may come apart, and networks of concepts may cease to work effectively or intelligibly. Simply replacing conceptual analysis with conceptual engineering risks preserving the same abstraction: treating concepts as definitions or application rules detached from their forms of life. My claim is that forms of life should form an explicit and systematic part of the methodology addressing conceptual disruptions.

The positive proposal of the paper is a situated conceptual methodology for addressing conceptual disruptions, which extends the object of analysis and engineering to include concepts within their larger ecosystems or forms of life. In the descriptive phase, *situated conceptual analysis* examines an existing concept's application rules alongside the background conditions that make them intelligible and effective, and diagnoses how changes in those conditions have unsettled the conceptual practice. In the normative phase, *situated conceptual ethics and engineering* imagine and evaluate possible continuations of the conceptual practice within possible forms of life. Addressing conceptual disruption in the face of rapidly evolving circumstances thus requires asking not only which concept we should adopt, but which combination of concept and conditions we have reason to bring about.

The paper is structured as follows. In section 1, I present the two main methodologies in the philosophy of AI: the descriptive-analytic methodology, centred on conceptual analysis, and the normative-engineering methodology, centred on conceptual engineering. In section 2, I present the open texture challenge to the descriptive-analytic methodology. In the last two sections, I present my proposal for a situated conceptual methodology to deal with conceptual disruptions, which includes situated conceptual analysis (section 3) and situated conceptual ethics and engineering (section 4).

1. Two methodologies for conceptual disruptions

1.1. Conceptual analysis and the descriptive-analytic methodology

Conceptual analysis is a family of methods whose primary goal is to identify or reconstruct the content of a concept or, on linguistic versions of the method, the meaning and conditions of use of a linguistic expression. When dealing with questions such as "is *a* F?", a methodology centred on conceptual analysis consists in using the conceptual analysis of the concept F, together with a characterisation of the phenomenon under investigation *a*, to answer the substantive philosophical question "is *a* F?". For example, using the methodology of conceptual analysis to answer the question "Are humans free?" consists in analysing the concept of freedom and characterising human action, in order to determine whether human actions fall under the concept. Only the first stage is conceptual analysis proper, although the term "conceptual analysis" is sometimes used to refer to the whole argumentative strategy. I will use "conceptual analysis" to refer to the first step, and "the descriptive-analytic methodology" to refer to the whole argumentative strategy.

Conceptual analysis can take many forms, along three related dimensions: the nature of its object, the means through which that object is investigated, and what form a successful analysis takes. Concepts can be understood as abstract objects or Fregean senses, psychological dispositions, patterns of linguistic usage, or inferential roles. This will in turn influence how their content is identified. The most classical form of conceptual analysis uses the "method of cases": thought experiments featuring different (combinations of) conditions, meant to elicit intuitions or judgments of competent speakers regarding the correct application of the concept or the correct use of the word (Horvath and Koch 2021). While this is usually done from the armchair, some have argued in favour of using experimental methods from psychology (Kölbel 2023). Finally, the resulting analysis can take the form of a strict definition in the form of biconditional with necessary and sufficient conditions, a biconditional with looser conditions such as "sufficiently resembles paradigms a-b-c", or a set of defeasible application norms. For the purposes of this paper, these differences do not matter; what matters is that conceptual analysis seeks to identify or reconstruct the content of a concept already operative in a linguistic community.

In its simplest form, the argumentative strategy has the following structure:

1. CONCEPTUAL PREMISE. According to analysis A, the application of the concept *F* requires or is otherwise governed by conditions C_1, C_2, …C_n.
2. CASE PREMISE. *a* satisfies (or fails to satisfy) conditions C_1, C_2, …C_n.
3. APPLICABILITY VERDICT. Therefore, F applies (or fails to apply) to *a*.

In the philosophy of AI, a significant number of contributions follow the descriptive-analytic strategy (Danaher 2019; Fröding and Peterson 2021; Munn and Weijers 2025; Lott and Hasselberger 2025; Gubelmann 2024; van Woudenberg et al. 2024; Ryan 2020; Carroll 2025; Hernandez Flowerman 2026; Dadà 2026; Brainard 2025; Runco 2023, among others). I will present two of them here. When asking "Can we be friends with an LLM?", Munn and Weijers (2025) argue that human-AI friendship is possible on the basis of a conception of friendship involving two requirements: 1) a preponderance of rewarding interactions; 2) mutual positive intentions between the parties involved. Based on this analysis of friendship, they argue that "human-AI friendships are possible because they [can] fulfil the criteria of rewarding interactions and shared positive intentions" (Munn and Weijers 2025). Similarly, in order to answer the question "Is ChatGPT an author?", van Woudenberg, Ranalli and Bracker first define an author as someone who engages, or has engaged, in "normal writing." Of their six features of normal writing, they regard intentionality about writing (the person who writes knows *that* they are writing), intentionality about the content of writing (the person who writes knows *what* they are writing), intentionality of the text (the text is about something), and ownership of its content as essential. They also offer a functional analysis of the concept of authorship, identifying five functions of authorship-ascriptions, which circumscribe the role that the concept of authorship currently plays in our linguistic practices. ChatGPT (in its current versions) is excluded as an author at these two levels. At the production level, ChatGPT outputs are not the products of normal writing and therefore lack authors (van Woudenberg et al. 2024, 12). At the attribution level, ChatGPT does not possess the agentic features that make the author-functions possible: although we can evaluate the text itself, we cannot normatively evaluate ChatGPT (ledger), hold it accountable (answerability), praise or blame it (desert), or demand a reparation (compensation), for what it wrote.

The conceptual premise is often inherited from previous philosophical theorising. For example, many rely on the Aristotelian definition of friendship in order to decide whether human-AI

friendships are currently possible or not, whether the verdict is positive (Danaher 2019) or negative (Fröding and Peterson 2021). Other times, the conceptual analysis is performed in the article itself: analysing how the concept is used in human cases, taking clear-cut cases and non-cases as their starting point, formulating a set of conditions capturing these cases, and testing it against further possible cases. For example, Munn and Weijers propose their own analysis of friendship (Weijers and Munn 2022; Munn and Weijers 2025) and refine it by considering objections to competing analyses that include additional necessary conditions, such as equality, voluntariness, or sentimentality. They argue that their analysis of friendship delivers the correct verdict on the same cases that are covered by competing analyses, such as slave-master relationships not counting as genuine friendships, and on additional cases for which the competing analyses deliver the wrong verdict, such as friendships between abled and disabled people or autistic people. This is a classical instance of the "method of possible cases", in which competing analyses are tested in scenarios for which there is a shared intuition regarding the correct verdict.

The case premise usually involves an empirical claim about the capabilities of present or near-future AI systems. More rarely, it involves a stronger modal or metaphysical claim that artificial systems could never possess a capacity such as consciousness, genuine intentionality or moral agency. The final applicability verdict is therefore supported by a combination of a conceptual premise and empirical or metaphysical premises.

The methodological strategy I am describing has two distinguishing features that will be relevant for the rest of this paper. First, its analysis takes as its starting point the concept as it operates in paradigmatic cases of human-centred phenomena, whether it is performed in the article itself or inherited from previous philosophical theory. Second, this analysis is used as authoritative in delivering a verdict for novel cases. Once the concept has been correctly analysed and the novel phenomenon adequately characterised, whether the concept applies to that phenomenon is taken to be settled. From this verdict, additional practical conclusions are often drawn. Munn and Weijers, for example, argue that "if [they] are right that human-AI bonds can count as genuine friendships, and those friendships can be valuable, there are implications … for how we should organize society" (Munn and Weijers 2025).

1.2. Conceptual engineering and the normative-engineering methodology

In recent years, philosophers have argued that conceptual analysis, which aims at identifying and revealing the content of our concepts, should be supplemented in our toolbox of philosophical methods by *conceptual ethics*, which aims at evaluating concepts according to some goals or desiderata, and *conceptual engineering*, which aims at creating or revising concepts in accordance with such goals or desiderata. This gives rise to a different argumentative strategy, normative rather than descriptive, centred on conceptual ethics/engineering rather than conceptual analysis. The *normative-engineering methodology* does not seek merely to determine whether an AI-related phenomenon falls under the concept we currently possess, but it asks which concept we should use or how our existing concept should be adapted to this novel context, given the purposes, values and practical consequences at stake. This is not to say that conceptual analysis disappears from the methodology (Díaz-León 2020). Analysing the concept currently operative may be essential for diagnosing issues or maintaining continuity of use. However, rather than being authoritative, the product of conceptual analysis constrains and informs the inquiry without determining its conclusion.

Here is the methodological procedure, in its most generic presentation, when concerned with a single concept F (rather than a broader conceptual repertoire):

1. CONCEPTUAL ANALYSIS: Analysis of the current concept F (and its conceptions F_1, F_2, F_3…).[1]
2. ETHICS:[2] Identification and justification of the conceptual functions, desiderata, values, practical stakes, relevant for evaluating the concept (in domain K).
3. CONCEPTUAL ETHICS: Evaluation of the concept F (or conceptions F_1, F_2, F_3…), and/or other conceptual alternatives, relative to the functions, desiderata, values, practical stakes identified in phase 2.

[1] Proponents of the normative methodology are more likely to recognize the multiplicity of existing conceptions without equating the concept's content with *one* of them.
[2] I follow Löhr (2024) in distinguishing "ethics" as a distinct phase from conceptual ethics and conceptual engineering (narrowly construed).

4. CONCEPTUAL ENGINEERING:[3] (if the evaluation of current concept in step 3 is negative) Creation of revised concept F', or selection of precisified concept F_n, requiring conditions C_a, C_b,…C_n, in accordance with the functions, desiderata, values, practical stakes identified in phase 2.
5. IMPLEMENTATION:[4] Adoption of the revised concept F' or the precisified concept F_n (in domain K).

The last two steps are added for conceptual engineering projects that seek to answer a question of the form "can F apply to *a*?", as is often the case in the philosophy of AI:

6. CASE PREMISE: *a* satisfies, or fails to satisfy, conditions C_a, C_b,…C_n.
7. APPLICATION VERDICT:[5] Therefore, F' or F_n applies (or does not apply) to *a*.

Köhler, Mecacci and Veluwenkamp's (2025) engineering of the concept of control provides a clear example of the normative-engineering methodology. Their project responds to AI-generated "responsibility gaps": because control is ordinarily treated as a necessary condition for attributing responsibility, autonomous systems appear to leave no human agent responsible. They start by identifying the dominant conception as that of *operational control*, understood as the capacity to intervene causally in a system's behaviour (conceptual analysis). They identify two valuable functions of the concept of control: locating agents who may appropriately bear responsibility and identifying those capable of preventing harm (ethics). They then argue that the conception *operational control* performs these functions poorly in the context of autonomous systems (conceptual ethics). They therefore recommend a conception of *meaningful human control* governed by two conditions: the system must track relevant human reasons, and its operation must be traceable to identifiable human agents who possess the appropriate capacities and understand their responsibilities (conceptual engineering). On this engineered conception, humans may control—and potentially be responsible for—even

[3] Interestingly, the "conceptual engineering" phase (narrowly construed) is optional, since the verdict in the previous step may be that the current concept should stay as it is. It is in this sense that Galvagni and Tollon (2026), who argue for preserving the existing concept of responsibility, consider that their project still falls under "conceptual engineering" broadly construed, that is to say, with the normative-engineering methodology.
[4] In conceptual engineering projects, this often stays at the level of a *recommendation*.
[5] Alternatively, the application verdict may become a criterion of evaluation, for example if we want our concept to apply (or not apply) to some AI phenomenon.

fully autonomous systems without intervening in their individual decisions (applicability verdict).

Given the continued prevalence of the descriptive-analytic strategy in the philosophy of AI, the superiority of conceptual engineering should not be taken for granted, and requires a positive argument. Why should we prefer the normative-engineering methodology to the descriptive-analytic methodology, when it comes to addressing conceptual questions raised by technological disruptions?

One way to defend this methodology is to point to the normative gap between the concept we possess and the concept we ought to possess:

> The actual content of concepts lacks the relevant normative significance to conclusively answer whether there can be AI responsibility gaps—or any other normative question for that matter. Even if we found evidence for what the actual content of our concept of responsibility is, it may yet not be the best content, the one that ought to be associated with the concept. (Himmelreich and Köhler 2022; similar argument in Viehoff 2023, 8)

A *descriptive* approach will fail to ground a *normative* conclusion; the analysis of our existing concept does not show that this concept is authoritative in establishing the way we should classify things or use the concept. This is a generic argument in favour of conceptual engineering; it does not explain why conceptual engineering is *especially* needed in cases of technological disruption.

But what if someone is not convinced by this argument, and believes the actual content of a concept *is* more authoritative than any of the constructed alternatives we may come up with? Three arguments suggested in the literature point to the superiority of the normative-engineering methodology *specifically* for conceptual disruptions.

The first appeals to the uncertainty or indeterminacy characteristic of conceptual disruption (Veluwenkamp et al. 2024; Löhr 2023a; Hopster and Löhr 2023; Dobler and Bracker 2026). New technologies generate cases for which our existing conceptual repertoire provides no uncontroversial classification. According to Himmelreich and Köhler, there is a "myriad alternatives that could precisify the concept" (*ibid.*, 14) or as Viehoff puts it, "we have options"

(2023, 8). The presence of alternatives means that analysis of the existing concept cannot by itself select among them. Conceptual engineering is better suited to such circumstances because it is concerned with conceptual decisions rather than conceptual discoveries: "When technological or social change leaves existing concepts underdetermined, the relevant question is often no longer which concept is correct, but which better serves the purposes of the practice in which it is employed" (Dobler and Bracker 2026, 7).

The second argument concerns adaptation to changing circumstances in the context of technological innovation. Even when our current concept or definition is clear and univocal, circumstances in which it was developed may have changed so substantially that we cannot assume that it remains adequate. Hence, conceptual engineering is "particularly sensible when social or technological advances pose problems for which our current conceptual repertoire might no longer be ideally suited" (Himmelreich and Köhler 2022, 3). This is why Hopster and Löhr (2023) characterize such conceptual engineering projects in the context of technological disruption as projects of *adaptation* rather than amelioration: their purpose is not necessarily to correct a concept that was always defective or ambiguous, but to "prevent our conceptual systems from degrading" (Hopster and Löhr 2023, 6).

The third argument concerns the practical and ethical stakes of conceptual decisions (Hopster and Löhr 2023; Löhr and Veluwenkamp 2025; Mollema 2026; Dobler and Bracker 2026). Whether an AI system is described as trustworthy, responsible, a potential friend, or an author can influence the expectations placed upon it, the responsibilities assigned to its users and developers, and the institutional arrangements and social norms constructed around it. As Juri Viehoff notes, conceptual engineering is especially urgent when there is ideological or commercial pressure to establish one rendering of a concept, for example, to frame AI systems as trustworthy in order to "de-emphasize practices … [of] *monitoring* and *strict regulatory oversight*" (Viehoff 2023, 9). In these contexts, a *normative* stance when dealing with conceptual questions appears appropriate in order to deal with the practical and ethical consequences of the use of an existing concept or the adoption of a new concept.

My aim is not to replace these arguments, but to provide a deeper account of why the first two hold. Why does technological novelty generate multiple available precisifications, and why can concepts that functioned adequately under earlier conditions cease to do so when those conditions change? In the next section, I provide an explanation by developing what I call the "open texture challenge" against the descriptive-analytic methodology.

2. Conceptual analysis and the open texture challenge

In this section, I argue that open texture presents a principled limit to conceptual analysis and defeats the descriptive-analytic strategy as a suitable methodology for resolving the conceptual questions raised by technological disruptions.

2.1. Two ways of understanding conceptual disruptions

The conceptual questions raised in the philosophy of AI arise because novel technologies are not only socially disruptive, but also *conceptually* disruptive. Here is Löhr's definition of "conceptual disruption":

> A conceptual disruption is any intentional or unintentional challenge or interruption of the ways in which an individual or group has intuitively classified individuals, properties, actions, situations, or events, leading to a classificatory conflict or uncertainty, i.e., a conflict or uncertainty about the application conditions of a word or concept. (Löhr 2023a, 3)

On this definition, the notion of conceptual disruption designates an observable condition: individuals or communities become uncertain or divided about how a concept applies. Conceptual questions in the philosophy of AI are clear examples of conceptual disruptions: there is extensive disagreement regarding how our concepts of responsibility, trust, agency, authorship should be applied to these novel phenomena. However, there are two ways of understanding conceptual disruptions. On an *epistemic* interpretation, conceptual disruptions are manifestations of a state of uncertainty: a concept possesses a determinate application, but we do not yet know what it is. The solution to conceptual disruptions is to *know* more: we need to discover what the content of the concept really is, or more information about the phenomenon under investigation, in order to determine how the concept should be applied in

these novel situations. For this way of understanding conceptual disruptions, the descriptive-analytic strategy makes sense. On a *semantic* interpretation, conceptual disruptions are signs of indeterminacy rather than uncertainty. We are "confront[ed] with a novel situation that our conceptual tools were not shaped to resolve" (Jorem and Löhr 2024, 946), which means that the answers to our questions cannot be found in the content of concepts that were shaped by and for very different conditions. In what follows, I will defend the second interpretation, by arguing that conceptual disruptions can be understood as manifestations of the *open texture* of concepts.

2.2. Open texture and forms of life

The notion of open texture was theorised by Waismann, a member of the Vienna Circle and close collaborator of Wittgenstein (Limbeck-Lilienau 2019), whose work certainly inspired that idea (Ben-Yami 2025). Waismann presents open texture as the character of concepts that are not "delimited in all possible directions" (Waismann 1945, 122). For such concepts, there is always the possibility of some "unforeseen factor" emerging that would create uncertainty or doubt with respect to their correct application, in a way that cannot be resolved by simply finding out more about either the concept or this novel factor.

In order to better understand open texture, it is useful to understand "open" or "incomplete" concepts in opposition to "closed" or "complete" concepts, in the context of Waismann's characterization of concepts as "thought models" or "rules of inference". On this understanding, a *closed* concept is "a thought model which anticipates and settles once for all every possible question of usage" (Waismann 1945, 125), that is to say, a finite set of rules that dictates the correct use of a concept and yields determinate verdicts in all possible situations. This is the case of formalised languages such as logic and mathematics, according to Waismann. By contrast, an *open* concept is a thought model which anticipates and settles many questions of usage, but not all of them: we do not have "rules ready for all imaginable possibilities" (Waismann 1945, 122). For example, in everyday situations, we can usually determine whether something is a cat, a tiger, or a dog. However, if we encountered a creature that looks and behaves like a cat, but suddenly grows "to a gigantic size" or was able to "be

revived from death", we would not know whether to say that "a new species has come into being" or that we found "a cat with extraordinary properties" (Waismann 1945, 121–22).

Where does the open texture of concepts come from? Waismann gestures at the "incompleteness of empirical description" (Waismann 1945, 126), which can never be exhaustive. I take open texture to be the character of concepts shaped by and embedded in practices and environments subject to unforeseen changes [redacted]. Open-textured concepts work effectively within a certain range of application because of stable background regularities. While they do often project effectively beyond the situations in which they are first used, changes in this stable background can cause our concepts to yield conflicting or indeterminate guidance. For example, our classification of biological species in terms of distinct evolutionary lineages works effectively because of regularities in the way animals reproduce and genetic traits are inherited. If interspecies breeding resulting in fertile offspring suddenly became widespread, our species concepts would yield indeterminate verdicts in a whole new range of cases. Technological changes introduce such disruptions in background regularities. Blackburn takes the example of the predicate "mother," "whose open texture is revealed [when] through technological advances differences open up between the mother that produces the ovum, the mother that carries the foetus to term, and the mother that rears the baby" (Blackburn 2008, under 'open texture'). Background regularities also include social norms and practices. In a society in which trans people are able to live in the gender they identify with, even the more precise concept "the mother who carries the foetus to term" can yield incoherent verdicts by both applying and not applying to the trans man who gave birth.

The dependence of concepts on forms of life is such that, even when concepts appear to have a determinate content, the disruption in background regularities may still make the conceptual practice unstable. In one of his examples of open texture, Waismann mentions the possibility of a piece of gold that emits "a new sort of radiation" (Waismann 1945, 122). While the classification of metals according to their atomic number is precise enough to warrant a determinate verdict in this imagined scenario, it would nevertheless become unstable if strange discoveries were repeatedly made with gold or other metals. After all, our current use of atomic number as the defining characteristic for metals stems from its ability to explain shared

properties and predict consistent behaviours. This point is important, because it explains why, in the case of technology-induced disruptions, even concepts that appear to have a determinate content or were given precise definitions in human contexts can become inadequate in novel conditions.

Put in Wittgensteinian terms, concepts are part of a *form of life* ("To imagine a language means to imagine a form of life", Wittgenstein 1953). Forms of life are the conditions and practices within which concepts work effectively and make sense, open texture is the result of this dependence of concepts on forms of life, and conceptual disruption is the mark of a change in a form of life that makes a concept partly break down.

Two remarks before I go on. First, Waismann's point is that open texture is a feature even of concepts defined in precise ways, such as in physical science. However, for the purposes of this paper, we do not even need to make that stronger claim. When it comes to the concepts involved in conceptual disruptions—love, friendship, agency, intentions, intelligence, responsibility, trust—most can agree that they are identified and articulated in overlapping webs of practices and whose normal use already includes a "penumbra of doubt" (Hart and Green 2012) beyond a stable core represented by paradigmatic cases. Second, my claim is not that open texture always leads to conceptual disruptions. On the contrary, we constantly make small decisions regarding how our concepts should be applied in novel or surprising situations without even realising it. Often, there is sufficient "agreement in form of life" or "in judgments" (Wittgenstein 1953, sects 241–242) to resolve the indeterminacy in some direction. Other times, we do not need to make such judgments, because we can treat the phenomenon as an exception and leave it unclassified, or we can manage the resulting conceptual disagreements. Conceptual disruptions arise when there is sustained uncertainty or division with respect to the way a conceptual practice should be continued or reorganised, and the consequences matter, either because the novel phenomena become widespread, or they are important to us, or both.[6]

2.3. The limits of conceptual analysis

[6] I thank Tamara Dobler for this point.

If we take seriously the idea that concepts are open-textured, and that conceptual disruptions are the results of significant disruptions in the forms of life that made our concepts effective, coherent, and valuable, what are the implications for conceptual analysis and its use to address such conceptual disruptions?

When it comes to the activity of conceptual analysis itself, open texture entails the limits of its scope: conceptual analysis cannot be *final* nor *exhaustive.* When made explicit in a definition or list of conditions, the content of an open-textured concept is best represented as including an implicit *ceteris paribus* clause or, as Waismann puts it, "the vague supposition that 'a normal situation subsists', that 'no disturbing factors are present' or in whatever way we may hint at the possibility of intervention of some unforeseen conditions" (Waismann 1945, 130). The popular "method of cases" in conceptual analysis consists in varying conditions in order to discover which conditions are genuinely relevant (or more strongly, necessary and sufficient) for the application of the concept. For example, when analysing the concept of personal identity, philosophers devise thought experiments in which the content of a mind is uploaded onto a computer, memories are swapped between two bodies, etc. (Noonan 2019). However, open texture means that we can reach boundary cases in which the concept breaks down, and instead of having clear intuitions regarding whether the concept *F* applies or not, we are uncertain or we disagree. (The fact that philosophers disagree about the correct verdict in some of these thought experiments may be an indication that they have reached those boundary cases.) Conceptual analysis can therefore produce conditions or rules of use that capture the concept's content or meaning within a certain range of application,[7] but it cannot produce an analysis guaranteed to remain stable across every possible context.

This has important consequences for the use of the descriptive-analytic methodology in conceptual disruptions. This argumentative strategy assumes that, once the content of a concept has been successfully analysed, that content can be used as an authoritative premise in determining how the concept should be applied to the novel phenomenon. In a conceptual

[7] Even then, the identification of the concept's content into a definite set of conditions is most often already a "sharpening" of its boundaries for a certain purpose (Dobler 2026).

disruption, however, the content revealed through conceptual analysis cannot simply be carried over in these extraordinary or novel cases to force a judgment in one particular direction. Because of the disruption in our form of life and the background regularities that previously made our concepts work effectively, the uncertainty we are faced with is not a mere matter of epistemic ignorance, but of semantic indeterminacy: there is no single correct way to continue the conceptual practice, as determined by the content of the concept, past usage, or the facts.

Why, then, does the descriptive-analytic strategy appear to yield determinate verdicts in the literature that we have cited? Let us have a closer look at these contributions.

Contributions using the descriptive-analytic strategy all take as their starting point paradigmatic examples of friendship, responsibility, or trust, occurring in human contexts.[8] However, generative AI technologies introduce significant changes in our environment, often by separating features that coincided in the human domain: the production of meaningful text and identifiable originators who can understand and stand behind it (Earp et al. 2026), the performance of a certain type of behaviour and the presence of subjective intentions and experiences, decision-making agency and moral agency. According to my account, because these features often or always travelled together, our concepts of authorship, friendship, or responsibility contain no instructions for cases in which they systematically come apart.

This is evidenced by the fact that the same paradigmatic cases can support different analyses of the concept, which are then used to argue for contradictory verdicts in novel cases. Van Woudenberg, Ranalli and Bracker analyse authorship through the conditions of "normal writing" and the functions of authorship-ascriptions. Although they allow an AI-generated text to acquire a human author through appropriation, they deny that ChatGPT itself is an author because it lacks proper intention and agency. By contrast, Polonsky and Rotman prioritise intellectual contribution in their analysis of academic authorship (Polonsky and Rotman 2023). Taking the Vancouver Protocol's four conditions as their starting point, they argue that ChatGPT can qualify as a co-author when it makes substantial intellectual and textual

[8] While there often already exist controversial cases for these concepts (e.g. can human-animal relationships count as friendships?), these cannot be used to support an analysis, precisely because they are not settled.

contributions, while interpreting approval and accountability functionally and collectively rather than as requiring the robust agential capacities demanded by van Woudenberg et al. Both accounts agree on paradigmatic human cases, in which contribution, ownership, and answerability ordinarily coincide; they disagree about whether ChatGPT should also count as an author once these features come apart. The same pattern appears in discussions of friendship, where some authors believe that what matters is rewarding interactions and positive intentions (Munn and Weijers 2025), while others believe that reciprocity and mutuality are fundamental (Elder 2017; Fröding and Peterson 2021).[9] In human cases, the question does not arise, because sincere goodwill, reciprocal attitudes, friend-like behaviour, shared activity and the capacity to benefit or suffer ordinarily occur together.

Accordingly, choosing between these proposals requires a precisification not determined by the antecedent concept alone. A feature is selected as essential, another is treated as merely contingent; a new priority rule is introduced for circumstances in which normally co-occurring features come apart. Each view preserves genuine continuity with the paradigmatic human cases while assigning priority to different features. What is presented as a neutral conceptual discovery is therefore already a conceptual decision. And what is presented as a disagreement about which philosopher has successfully discovered the concept's hidden extension is therefore best understood as a practical disagreement concerning two ways of continuing the conceptual practice.

Open texture provides a distinctive challenge that directly undermines the descriptive-analytic methodology in the case of conceptual disruptions, even for those who do not believe in the automatic superiority of conceptual engineering over conceptual analysis. At this stage, we could turn directly to the normative-engineering methodology described in section 1.2. However, as I explain in the next section, the idea of open texture provides another methodological lesson, one that is not systematically taken into account in the conceptual engineering literature.

[9] Sometimes the same analysis (in appearance) can be used to deliver contradictory verdicts, because of how the components of the analysis are themselves interpreted or determined; for example, the Aristotelian definition of friendship was also used to ground the claim that we can be friends with LLMs (Danaher 2019).

3. Situated conceptual methodology: the descriptive phase

So far, I have shown that open texture presents a direct challenge to the use of conceptual analysis in grounding a verdict in the conceptual questions raised by AI. However, this is not the only methodological lesson that can be drawn from my argument. Open texture reminds us not only that concepts are always potentially indeterminate, but also that they do not live in vacuums. Addressing conceptual disruptions therefore requires paying attention not only to existing or candidate concepts, but also to the environments on which they depend to function well. In this section and the next, I defend my proposal for a *situated* version of the normative-engineering methodology, in which the object of conceptual analysis, ethics and engineering is extended to include forms of life. Although this integration is often done partly and implicitly in the existing literature, for example through an understanding of the concept's functions or an examination of the practical consequences of concepts, my situated methodology proposes to do this systematically and explicitly.

3.1. Situated conceptual analysis

To address a conceptual disruption, the descriptive phase of the proposed methodology must determine *what* worked effectively in our previous conceptual practice that no longer works now, and *why* it no longer works now. For this, it is not sufficient to extract or identify the content of the concept, henceforth "narrow conceptual analysis". We must also know what background conditions made that concept effective and intelligible, which are now no longer in place. This is the task of *situated* conceptual analysis, which looks at an existing concept (or network of concepts) and the forms of life that make its use effective, coherent, valuable. For now, let's say that concepts are effective when they generally yield determinate verdicts that allow us to make further inferences; coherent when they do not generally produce incoherent or conflicting verdicts, either directly, in their immediate application or indirectly, in combination with other concepts; valuable when they allow us to fulfil our goals or allow practices to function well.

How do we include forms of life—a notion that Wittgenstein intentionally left vague—into the scope of our analysis? To parse out this notion, I use Queloz's notion of *need matrix* (Queloz 2025) which involves a) circumstances (including features of our natural, social, technological environment), b) capacities (including natural, social, technologically-enhanced), c) concerns (including our values, purposes, projects). To illustrate his need matrix, Queloz takes the example of health. We have a concern in staying alive and functional. But "if our physiological condition never changed [i.e. different capacities], thinking in terms of the concept of health would be pointless for us" (Queloz 2025, 223). The concept of health would also not make sense to us—except imaginatively, if we could imagine a kind of people whose physiological condition did change. While this is the analytical tool I propose for understanding forms of life, other forms of situated conceptual analysis may use different tools.

The first step of situated conceptual analysis is to make the unit of analysis explicit. For example, the concept of authorship does not have the same content nor functions in the same way in the bookselling business, in academic publishing in the humanities, or in academic publishing in the natural sciences, where different features will be foregrounded: the one who produces the prose, the one who has the ideas, the one who shapes the direction of the paper or project, the one who stands behind the claims being made (Earp et al. 2026).[10] Selecting a unit of analysis is rarely neutral — which forms of life we focus on depends on our projects, but may also rest on unexamined assumptions of what counts as "normal" conditions.

The next step is to analyse the practical dependence between concepts and forms of life. How so? Usually, the presence of a conceptual disruption provides a clue as to the elements of a form of life that contributed to the concept's working effectively, but are no longer present. Additionally, the traditional method of cases used in classical forms of conceptual analysis can be repurposed to identify patterns of failures that map onto different components of a form of life. By testing our judgments in different real or imaginary scenarios, we can uncover conditions in which the concept yields indeterminate or incoherent verdicts. Instead of asking "What is the correct verdict in this case?" (when there is none), situated conceptual analysis

[10] I do not take a stand regarding the individuation criteria for concepts, whether we should talk of conceptions rather than concepts, or distinct but interrelated concepts expressed by the term 'author'.

asks: “What explains our hesitation, disagreement or attraction to competing verdicts?”. This is possible because our uncertainty or disagreement is usually *contentful*: we don’t just say “I don’t know whether F applies”, we also say, “because on the one hand, condition a is present, but on the other hand, condition b is absent. *Usually*, conditions a and b come together”. These kinds of thought experiments can help us determine whether some circumstance, capacity, or concern is a key component of a form of life for a specific concept or network of concepts, and which are merely contingent features.[11]

However, situated conceptual analysis is best practised by combining armchair methods with empirical sources from history, anthropology, sociology providing information on the way concepts were used and evolved in different forms of life. Studying conceptual practices from the past or from other communities can reveal the contingency of features found in our own practices that might otherwise appear as necessary or universal (Dutilh Novaes 2015).

3.2. Comparison with functional analyses: the case of authorship

How does situated conceptual analysis differ from functional analyses of a concept, which look not only at the content of the concept, but more broadly at the functions that the concept fulfils in our practices? This section will use van Woudenberg et al.’s functional analysis of authorship as a contrast to illustrate the specificity of situated conceptual analysis.

Van Woudenberg et al. (2024) define authorship in relation to what they call “normal writing”, and identify five functions of authorship ascriptions, including keeping a ledger of contributions, assigning answerability, and distributing praise and blame. This situates the concept within a wider context of social and normative practices in which we judge each other, hold each other accountable, and reward each other for what we write. But identifying these functions does not yet identify the *conditions* that enable the successful performance of these functions. A situated conceptual analysis would go further and identify the circumstances, capacities, and concerns that sustain the work that a concept does for us.

[11] Situated conceptual analysis does not aim to catalogue every background condition on which a conceptual practice depends. Such a catalogue would be indefinitely extensive, encompassing conditions such as gravity or the persistence of objects. The inquiry selectively identifies conditions whose alteration is relevant to actual or foreseeable disruptions.

The functions performed by the concept of authorship presuppose not only that we care about who is answerable or who deserves praise (concerns), but that there is usually *someone* who can be answerable or praised for the production of a text. Hence, these functions can be reliably performed under the expectation that texts are usually written by individuals who intentionally write and understand what they wrote (circumstances) and who are capable of taking responsibility and ownership of their writing (capacities)—in short, conditions in which "normal writing" *is the norm*. Thus, under "normal" conditions, uncertainty regarding authorship ascriptions concerns the identity of the text originators (for example when investigating the source of anonymous medieval poems) or their level of contribution to the text (for example when assessing plagiarism or co-authorship).

By contrast, when a text is written in a trance, hypnosis, or automatic writing, the concept of authorship cannot perform its functions—we are unable to clearly identify the originator of the text as properly responsible for its content. These are "abnormal" conditions: as van Woudenberg et al. note, these cases are "clear exceptions to what normally is the case when one is writing" (2024, 6). Here, the uncertainty concerns not just the question "who is the author?" but whether or how the concept of authorship even applies. Precisely because these cases are "exceptions", the concept of authorship (until the arrival of LLMs) could reliably perform its functions.

With the rise of sophisticated LLMs, we have novel conditions in which textual production can be systematically separated from sentient agents intentionally writing texts and being accountable for them. As a result, the concept of authorship becomes less *valuable*, or cannot properly fulfil its functions. If we *do* apply the concept to ChatGPT, the concept cannot fulfil many of its normative functions, as noted by van Woudenberg et al. If we *don't* apply the concept to ChatGPT, but to a human proxy, the concept also fails to fulfil its normative functions in a satisfying way, since not crediting LLMs as authors leads to an overinflation of credit for humans whose research and/or writing are mostly conducted by LLMs (Hurshman et al. 2026). If we don't apply the concept to ChatGPT nor to human proxies but conclude that these texts are authorless, we end up with a mass of authorless texts, which again prevents us

from assigning responsibility for the claims made in the text, one of the main uses that make authorship ascriptions valuable for us. It may well be that, under these novel conditions, the different jobs originally performed by our concept of authorship "pull apart" and can no longer be satisfied by a single concept (Earp et al. 2026).

Instead of using the disruption of such conditions to ground a negative application verdict (ChatGPT is *not* an author), situated conceptual analysis highlights the disruption of the form of life that made the concept of authorship previously work well. This means that something needs to change in our conceptual practice to resolve those indeterminacies, incoherences, and functional breakdowns.

4. Situated conceptual methodology: the normative phase

We now move to the normative phase of the methodology, which is concerned with the direction that our conceptual practices should take in the face of change and disruption. Like conceptual analysis, I argue that the scope of conceptual ethics and engineering needs to be extended to concepts *in their forms of life*. Many conceptual engineers already defend the view that concepts are embedded in social practices, according to which evaluating and intervening on concepts is inseparable from evaluating and intervening on social practices (Haslanger 2012; Podosky 2026). My view of situated conceptual engineering is in line with those proposals, and specifically looks at how this is done when addressing conceptual disruptions.

4.1. From functional conceptual engineering to situated conceptual engineering

Conceptual disruptions bring home the fact that the functions of concepts cannot be separated from the environments in which they operate. A concept that functioned effectively under one set of conditions may no longer function well under new conditions. As we have seen in the case of authorship, preserving the original concept in its strictest boundaries (by excluding LLMs as candidate authors) in view of preserving its functions in no way guarantees that the concept will be able to satisfy them adequately in the novel conditions. A functional approach to conceptual engineering, in this context, takes the form of "conceptual adaptation", where

the aim is to "[preserve] conceptual functions to adapt to new circumstances" (Hopster and Löhr 2023, 3). For this, conceptual functions must be understood in relation to the environments and practices that enable their performance, rather than solely in terms of intended functions. The notion of conceptual *affordances*, borrowed from ecological psychology, makes this relational character explicit: what a concept allows us to do is always "relative to a user and a given context" (Marchiori 2025, 3). Queloz's framework similarly recasts functional evaluation in terms of *conceptual needs* arising from a combination of concerns, capacities, and circumstances (Queloz 2025).

This ecological turn is a first step towards situated conceptual engineering. However, it does not fully capture the interdependence between concepts and forms of life, in two respects. First, the goal of "preserving conceptual functions" assumes that the work that a concept did for us, in the pre-disrupted form of life, *should be the work that the concept does for us*, in the new form of life. However, a disruption may change our form of life in such a way as to generate new concerns, or weaken our commitment to inherited ones. In the case of authorship, for example, we cannot take for granted that preserving the concept's original functions should be our priority. The disruption caused by LLMs could also be seen as an opportunity for reorganising academic practices into "hypercommons", such that "individual authorial credit is less important than collective knowledge production" (Islam and Greenwood 2024, 660). The priority given to *preserving* functions or old concerns is already a normative choice, which should itself be made clear or justified. As Queloz notes, we should not take familiar concerns for granted, but "critically identify" with them, upon further reflection (Queloz 2025, 23). Situated conceptual engineering remains open to the possibility that changing forms of life alter not only what our concepts do for us, but also what we want them to do.

There is a second limitation. Not only does the idea of conceptual adaptation hold fixed the original functions of the concept, but also the new conditions in which it operates. But just as original functions should not automatically be preserved, newly emerging conditions should not automatically be accommodated. Even Queloz, who emphasizes the malleability of concerns, suggests that once a needs matrix has been identified, it works as a unilateral constraint on conceptual choice: "Like the matrix from which printing type is cast, the need

matrix acts as a mould from which to cast fitting [concepts]" (Queloz 2025, 22). But forms of life are not fixed, they can themselves become the object of critical evaluation and intervention: they include concerns or values that we can uphold or dismiss, circumstances that we can act upon, capacities that we can change (within limits, of course). This is all the more true in the kinds of conceptual disruptions we are interested in, which are caused and accompanied by rapid changes in technological developments, institutional structures, and social practices.

Furthermore, choosing a conceptual practice often amounts to choosing a form of life, where adopting a new concept either requires and/or fosters concomitant changes in attitudes, practices, institutions, technologies. This is the case for all concepts, but even more so when the new form of life has not stabilised yet. As Dobler and Bracker put it, "Rather than merely adapting inherited concepts to technological change, judges, regulators, and legislators actively shape how new technologies are understood and governed" (Dobler and Bracker 2026, 8). Thus, situated conceptual engineering does not only ask which concept would best fit existing conditions, but which combination of concept and form of life we have reason to bring about. The question of "what concept do we want?" becomes inseparable from the question "what society do we want to live in?".

4.2. Situated conceptual ethics and engineering: a more holistic outlook

In the normative phase of the situated methodology, the object of evaluation and engineering is not concepts in isolation but the relevant *concept + form-of-life* configurations. The methodology draws on Queloz's framework (2025), while integrating the remarks just made. The presentation of the methodology is programmatic, rather than providing detailed procedures to follow.

Situated conceptual analysis. As I have explained in the previous section, this phase provides an account of how the earlier conceptual practice worked, the conditions that enabled it to work, and a diagnosis of what changed and unsettled the conceptual practice. However, this diagnosis does not turn the previous functions into a benchmark, nor emerging form of life into a new set of givens to which the concept must adapt. Instead, it opens a broader normative

inquiry into which elements of the emerging form of life can and should be preserved, revised, or transformed.

Situated ethics. Queloz's notion of "needs" provides a more situated criterion of evaluation than isolated concerns, purposes, or values, by anchoring them in circumstances and capacities. My proposal further emphasizes their *dynamic* character: new concerns may emerge, or old concerns may shift, in their interaction with new circumstances and capacities; conversely, circumstances and capacities are continuously shaped by our concerns. This means that, especially when it comes to conceptual disruptions, we rarely have a stable "need matrix" to use as a criterion of evaluation for our concepts, but rather, changing conditions that have not stabilised yet into a new form of life. For example, the current crisis in authorship comes from the fact that what used to be "normal writing" is rapidly ceasing to be the norm, but *the new normal has not yet settled.* Accordingly, situated ethics is about finding a general direction of change that aligns with both what we determine to be feasible and what we find valuable.

Situated conceptual ethics and engineering thus become part of a larger project evaluating and shaping forms of life. The larger project asks: "What form of life do we want? What should the new normal be?" Situated conceptual ethics and engineering ask: "For each of these possibilities, what concept or network of concepts would work effectively, coherently, and valuably?" Thus, the proper object of conceptual ethics and engineering is the *configuration of concept and form of life* that has a good internal fit (internal evaluation) and that we want to bring about (external evaluation). Choosing how to continue the conceptual practice of authorship cannot be done without choosing a way of arranging our research, creative, journalistic practices (even staying with the current organisation is a choice).

4.3. A collective and interdisciplinary methodology

Situated conceptual engineering is best done collectively. We are deliberating on what kind of society "we" want to live in, which is a matter of public and pluralistic debate. Public, because those affected by and participating in a conceptual practice possess relevant experience, interests and knowledge, and should participate in determining which possible configurations

are entertained. Pluralistic, because evaluating forms of life or new normals will involve conflicting perspectives about concerns and values.

The project must also be interdisciplinary, not least because the feasibility and consequences of possible concept + forms of life configurations cannot be established from the armchair. Engineers can suggest which technological capacities are possible and alterable. Legal and policy experts can assess institutional possibilities and constraints. More importantly, at the implementation level, conceptual practices cannot be changed independently of the legal, institutional and organisational structures through which they are realised. Philosophers still retain a distinctive role in articulating conceptual alternatives and examining how well they fit the possible forms of life under consideration, even if there as well, artists and science fiction writers can help us envision unfamiliar forms of life and ways of speaking. Philosophers' conceptual work is therefore situated within, and informed by, a broader collaborative inquiry.

My last point concerns the issue of implementation. Changing concepts is already difficult, now we need to change forms of life too? I will give two brief responses. First, as I have already pointed out, in conceptual disruptions, things are already changing; we may be able to intervene and steer the direction of change. Second, the goal of situated conceptual engineering is not (always) to *resolve* the conceptual disruption in one way or another, but to *negotiate* the direction of the change. In some cases, a resolution needs to happen (e.g. responsibility understood in a legal sense, authorship attributions in academia). In other cases, such as friendship vs mere companionship, trustworthy AI vs reliable AI paradigms, the contestation will likely go on, precisely because competing visions of society are at stake. This methodology is meant to help articulate and defend proposals for new conceptual practices and identify where to focus efforts for change.

Conclusion

Conceptual disruptions result from concomitant disruptions in the form of life that previously sustained the conceptual practice, leaving us without an adequate vocabulary to describe what is happening, or pulling us in different directions. Instead of trying to find the answers in the

"core" of how our concepts have been used so far, my proposed methodology proceeds as follows. First, situated conceptual analysis maps the concept in relation to its form of life in order in order to identify the source of the conceptual disruption. Because nothing in the concept itself, or the empirical facts, predetermines how the conceptual practice should be continued, a normative and constructive approach is needed, rather than a merely descriptive one. Second, situated conceptual ethics and engineering will evaluate and imagine possibilities for the future, as different configurations of concepts and forms of life. The task is therefore not merely to adapt our concepts to an emerging new normal, but to ask which new normal our conceptual practices should help bring about.

References

Ben-Yami, Hanoch. 2025. 'Wittgenstein and Waismann's Open Texture'. *Philosophical Investigations* 48 (3): 318–32. https://doi.org/10.1111/phin.12454.

Blackburn, Simon. 2008. 'Open Texture'. In *The Oxford Dictionary of Philosophy*. Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780199541430.001.0001/acref-9780199541430.

Brainard, Lindsay. 2025. 'The Curious Case of Uncurious Creation'. *Inquiry* 68 (4): 1133–63. https://doi.org/10.1080/0020174X.2023.2261503.

Carroll, Nicholas George. 2025. 'Artificial Goodwill and Human Vulnerability: The Case for Building Merely Reliable, Rather than Trustworthy, Artificial Intelligence Technologies'. *Philosophy & Technology* 38 (2): 65. https://doi.org/10.1007/s13347-025-00881-w.

Dadà, Silvia. 2026. 'Can AI Care for Patients? A Conceptual Incompatibility'. In *The Palgrave Handbook on the Ethics of Artificial Intelligence*, edited by Steven S. Gouveia. Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-15112-4_26.

Danaher, John. 2019. 'The Philosophical Case for Robot Friendship'. *Journal of Posthuman Studies* 3 (1): 5–24. https://doi.org/10.5325/jpoststud.3.1.0005.

Díaz-León, Esa. 2020. 'Descriptive vs. Ameliorative Projects: The Role of Normative Considerations'. In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett. Oxford University Press. https://doi.org/10.1093/oso/9780198801856.003.0009.

Dobler, Tamara. 2026. ‘Noam Chomsky’s Wittgenstein: On Things (Not) to Do to Concepts’. In *The Routledge Handbook of Wittgenstein and Other Philosophers: Part 1*. Routledge.

Dobler, Tamara, and Daniel Bracker. 2026. ‘Should the Legal Concept of Speech Include Chatbot Output in US Law?’ *Philosophy & Technology* 39 (3): 174. https://doi.org/10.1007/s13347-026-01187-1.

Dutilh Novaes, Catarina. 2015. ‘Conceptual Genealogy for Analytic Philosophy’. In *Beyond the Analytic-Continental Divide*, edited by Jeffrey A. Bell, Andrew Culrofello, and Paul M. Livingston. Routledge. https://doi.org/10.4324/9781315766621-8.

Earp, Brian D., Sebastian Porsdam Mann, Tsutomu Sawai, and Tenzin Wangmo. 2026. ‘Death, Authorship and Generative AI: A Call for Commentaries’. *JME Practical Bioethics* 2 (2). https://doi.org/10.1136/jmepb-2026-000263.

Elder, Alexis M. 2017. *Friendship, Robots, and Social Media: False Friends and Second Selves*. Routledge.

Fröding, Barbro, and Martin Peterson. 2021. ‘Friendly AI’. *Ethics and Information Technology* 23 (3): 207–14. https://doi.org/10.1007/s10676-020-09556-w.

Galvagni, Enrico, and Fabio Tollon. 2026. ‘Engineering Responsibility in the Age of AI: Amelioration or Preservation?’ *Inquiry* 0 (0): 1–19. https://doi.org/10.1080/0020174X.2026.2615772.

Gubelmann, Reto. 2024. ‘Large Language Models, Agency, and Why Speech Acts Are Beyond Them (For Now) – A Kantian-Cum-Pragmatist Case’. *Philosophy & Technology* 37 (1): 32. https://doi.org/10.1007/s13347-024-00696-1.

Hart, H. L. A., and Leslie Green. 2012. *The Concept of Law*. 3rd edition. Edited by Joseph Raz and Penelope A. Bulloch. Oxford University Press.

Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199892631.001.0001.

Hernandez Flowerman, Camila. 2026. ‘AI, LLMs, and the Normativity of Belief’. *Synthese* 207 (4): 132. https://doi.org/10.1007/s11229-026-05475-3.

Himmelreich, Johannes, and Sebastian Köhler. 2022. ‘Responsible AI Through Conceptual Engineering’. *Philosophy & Technology* 35 (3): 60. https://doi.org/10.1007/s13347-022-00542-2.

Hopster, Jeroen, and Guido Löhr. 2023. ‘Conceptual Engineering and Philosophy of Technology: Amelioration or Adaptation?’ *Philosophy & Technology* 36 (4): 70. https://doi.org/10.1007/s13347-023-00670-3.

Horvath, Joachim, and Steffen Koch. 2021. ‘Experimental Philosophy and the Method of Cases’. *Philosophy Compass* 16 (1): e12716. https://doi.org/10.1111/phc3.12716.

Hurshman, Clint, Sebastian Mann, Julian Savulescu, and Brian Earp. 2026. *LLMs as Co-Authors: The Argument from Fairness*.

Islam, Gazi, and Michelle Greenwood. 2024. ‘Generative Artificial Intelligence as Hypercommons: Ethics of Authorship and Ownership’. *Journal of Business Ethics* 192 (4): 659–63. https://doi.org/10.1007/s10551-024-05741-9.

Jorem, Sigurd, and Guido Löhr. 2024. ‘Inferentialist Conceptual Engineering’. *Inquiry* 67 (3): 932–53. https://doi.org/10.1080/0020174X.2022.2062045.

Köhler, Sebastian, Giulio Mecacci, and Herman Veluwenkamp. 2025. ‘Responsibly Engineering Control’. *American Philosophical Quarterly* 62 (2): 113–32. https://doi.org/10.5406/21521123.62.2.02.

Kölbel, Max. 2023. ‘Varieties of Conceptual Analysis’. *Analytic Philosophy* 64 (1): 20–38. https://doi.org/10.1111/phib.12249.

Limbeck-Lilienau, Christoph. 2019. ‘Waismann in the Vienna Circle’. In *Friedrich Waismann: The Open Texture of Analytic Philosophy*, edited by Dejan Makovec and Stewart Shapiro. Springer International Publishing. https://doi.org/10.1007/978-3-030-25008-9_2.

Löhr, Guido. 2023a. ‘Conceptual Disruption and 21st Century Technologies: A Framework’. *Technology in Society* 74 (August): 102327. https://doi.org/10.1016/j.techsoc.2023.102327.

Löhr, Guido. 2023b. ‘If Conceptual Engineering Is a New Method in the Ethics of AI, What Method Is It Exactly?’ *AI and Ethics*, ahead of print, May 16. https://doi.org/10.1007/s43681-023-00295-4.

Löhr, Guido. 2024. ‘What’s the Relation between Conceptual Ethics and Ethics? An Instrumentalist Defense of Conceptual Engineering’. In *Conceptual Engineering*. Brill mentis. https://doi.org/10.30965/9783969753026_009.

Löhr, Guido, and Herman Veluwenkamp. 2025. ‘Rethinking Philosophical Methodology: Conceptual Engineering Meets Value Sensitive Design’. *Metaphilosophy* 56 (3–4): 328–43. https://doi.org/10.1111/meta.12730.

Lott, Micah, and William Hasselberger. 2025. ‘With Friends Like These: Love and Friendship with AI Agents’. *Topoi*, ahead of print, August 31. https://doi.org/10.1007/s11245-025-10247-8.

Marchiori, Samuela. 2025. ‘Conceptual Affordances: (How) Should They Inform Conceptual Engineering?’ *Synthese* 206 (2): 88. https://doi.org/10.1007/s11229-025-05161-w.

Mollema, Warmhold Jan Thomas. 2026. ‘The Apt Curation Model and the Need for More Conceptual Ethics of AI and Authorship’. *Philosophy & Technology* 39 (1): 41. https://doi.org/10.1007/s13347-026-01050-3.

Munn, Nick, and Dan Weijers. 2025. ‘Human–AI Friendship Is Possible and Can Be Good’. In *Oxford Intersections: AI in Society*, 1st edn, edited by Philipp Hacker. Oxford University PressOxford. https://doi.org/10.1093/9780198945215.003.0076.

Noonan, Harold W. 2019. *Personal Identity*. Third Edition. Routledge.

Podosky, Paul-Mikhail Catapang. 2026. ‘Conceptual Economics *Contra* Conceptual Engineering’. *Inquiry*, February 16, 1–19. https://doi.org/10.1080/0020174X.2026.2617389.

Polonsky, Michael Jay, and Jeffrey D. Rotman. 2023. ‘Should Artificial Intelligent Agents Be Your Co-Author? Arguments in Favour, Informed by ChatGPT’. *Australasian Marketing Journal* 31 (2): 91–96. https://doi.org/10.1177/14413582231167882.

Queloz, Matthieu. 2025. *The Ethics of Conceptualization: Tailoring Thought and Language to Need*. Oxford University Press.

Runco, Mark A. 2023. ‘AI Can Only Produce Artificial Creativity’. *Journal of Creativity* 33 (3): 100063. https://doi.org/10.1016/j.yjoc.2023.100063.

Ryan, Mark. 2020. ‘In AI We Trust: Ethics, Artificial Intelligence, and Reliability’. *Science and Engineering Ethics* 26 (5): 2749–67. https://doi.org/10.1007/s11948-020-00228-y.

Tanswell, Fenner Stanley. 2018. ‘Conceptual Engineering for Mathematical Concepts’. *Inquiry* 61 (8): 881–913. https://doi.org/10.1080/0020174X.2017.1385526.

Veluwenkamp, Herman, Jeroen Hopster, Sebastian Köhler, and Guido Löhr. 2024. ‘Socially Disruptive Technologies and Conceptual Engineering’. *Ethics and Information Technology* 26 (4): 65. https://doi.org/10.1007/s10676-024-09804-3.

Viehoff, Juri. 2023. ‘Making Trust Safe for AI? Non-Agential Trust as a Conceptual Engineering Problem’. *Philosophy & Technology* 36 (4): 64. https://doi.org/10.1007/s13347-023-00664-1.

Waismann, Friedrich. 1945. ‘Verifiability’. *Proceedings of the Aristotelian Society, Supplementary Volumes* 19: 119–50.

Weijers, Dan, and Nick Munn. 2022. ‘Human-AI Friendship: Rejecting the Appropriate Sentimentality Criterion’. In *Philosophy and Theory of Artificial Intelligence 2021*, edited by Vincent C. Müller. Springer International Publishing. https://doi.org/10.1007/978-3-031-09153-7_17.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Wiley-Blackwell.

Woudenberg, René van, Chris Ranalli, and Daniel Bracker. 2024. ‘Authorship and ChatGPT: A Conservative View’. *Philosophy & Technology* 37 (1): 34. https://doi.org/10.1007/s13347-024-00715-1.