Meaning, Form and the Limits of Natural Language Processing

Dürr, Oliver ; Segessenmann, Jan ; Steinmann, Jan Juhani

DOI: https://doi.org/10.1628/ptsc-2023-0005

Oliver Dürr, Jan Segessenmann and Jan Juhani Steinmann

Meaning, Form and the Limits of Natural Language Processing[1]

This article engages the anthropological assumptions underlying the apprehensions and promises associated with language in artificial intelligence (AI). First, we present the contours of two rivalling paradigms for assessing artificial language generation: a holistic-enactivist theory of language and an informational theory of language. We then introduce two language generation models – one presently in use and one more speculative: Firstly, the transformer architecture as used in current large language models, such as the GPT-series, and secondly, a model for 'autonomous machine intelligence' recently proposed by Yann LeCun, which involves not only language but a sensory-motor interaction with the world. We then assess the language capacity of these models from the perspectives of the two rivalling language paradigms. Taking a holistic-enactivist stance, we then argue that there is currently no reason to assume a human-comparable language capacity in LLMs and, further, that LeCun's proposed model does not represent a significant step toward artificially generating human language because it still lacks essential features that underlie the linguistic capacity of humans. Finally, we suggest that proponents of these rivalling interpretations of LLMs should enter into a constructive dialogue and that this dialogue should continuously involve further empirical, conceptual, and theoretical research.

Keywords: AI, Large Language Models (LLMs), Language, Philosophy of language, Meaning, Understanding, Common-sense, Enactivism, Information, Artificial general intelligence (AGI), Autonomous machine intelligence, Embodied cognition

1 We are grateful to the *Institute of Hermeneutics and Philosophy of Religion* at the University of Zurich, the *Center for Faith & Society* at the University of Fribourg, the *Institut Catholique de Paris*, as well as *Harris Manchester College*, Oxford for providing us with the resources and support for writing this paper.

PTSc 10 (2023), 42–72 DOI 10.1628/ptsc-2023-0005
ISSN 2195-9773

1. Introduction

Language remains in many ways
a mysterious thing.
(Taylor 2016, 341)

Artificial intelligence (AI) has recently reached new levels with generative AI systems. State-of-the-art 'large language models' (LLMs), for example, perform a variety of tasks incredibly well. They generate written text in reaction to prompts from human interlocutors that human beings perceive as coherent, meaningful, and often helpful. The practical capabilities of LLMs in many fields of application are only just being discovered, and the technology stands to transform our societies fundamentally.

These linguistic outputs of LLMs are of particular interest to philosophical and theological anthropology because, at least since Aristotle's seminal definition of human beings as ζῷον λογικόν (or ζῷον λόγον ἔχον), we have considered ourselves to be *uniquely* gifted with linguistically mediated reason[2]. Language, for humans, is not just a tool we acquire and deploy to communicate but rather the medium through which we inhabit the world and make sense of things and, particularly, our lives. Language is "a feature of what we are" (Taylor 2016, 90). It has become part of the definition of what it means to be human; therefore, any definition or theory of language is always also bound up with a specific anthropology. Current breakthroughs in the linguistic performance of LLM systems thus raise essential questions about the nature of language, technology, and human beings.

One apparent question arising from the human-like output of LLMs is whether this indicates mental capacities (or models of such capacities) in these systems. In some sense, LLMs have outperformed critics, who would classically have argued that the production of meaningful text depends upon insight and understanding of the meaning of text and that AI, not capable of understanding, would run into serious limitations when faced with the task of generating meaningful text (or at least: text that human beings can meaningfully interpret) (see, e. g., Dreyfus 1992, Dreyfus 2012). Those critiques require an update (Durt, Froese, and Fuchs 2023). Alan Turing,

2 Aristotle's definition is usually translated via the Latin (*animal rationale*) as 'rational animal.' Charles Taylor, however, helpfully proposes to render this as "animal possessing 'logos'" whereby '*logos*,' depending on the context, can mean 'word,' 'discourse,' or 'account' among others. This leads to Taylor's rendering as "animal possessing language" or "language animal" (Taylor 2016, 338). Some traditions in theological anthropology have also positively taken up Aristotle's definition, and most of them see human beings as participating in some form or other in the divine λόγος.

who, by contrast, seemed convinced that machines would be capable of this, proposed the 'Turing test' (testing if a system can generate convincingly meaningful text) as a benchmark for measuring if a machine can really 'think' (Turing 1950)[3]. Current LLMs are approaching this level of conversational plausibility.

However, similar behaviour can emerge from vastly different processes (Searle 1999; Neisser 1967), and it is not clear, philosophically, how this fact must be interpreted: It could be used to stress the *similarity* of computational processes and human cognition, e.g., in mind-body functionalism (Putnam 1975), a computational theory of mind (Chalmers 2011; see Rescorla 2020), or indirectly via the notion of 'information-processing' (Neisser 1967; Tegmark 2017). All of those interpretations rest upon the assumption of 'substrate independence' or 'multiple realisation' of mind. Precisely this assumption, however, has increasingly been called into question, not least in light of the role organic life has as a basis for cognition (Fuchs 2022; Polger and Shapiro 2016; see sections 2.a and 3.b). Others argue that we are dealing here with nothing more than 'stochastic parrots' (Bender et al. 2021), which will not emerge as sentient agents. Rather, they present a host of other challenges to our societies – ecological, economic, and political in nature (Crawford 2021), and we would add: cultural as well as spiritual (Hoff 2021; Dürr 2021). Suppose, however, one argues for the *dissimilarity* of computational processes and human cognition. In that case, however, the outputs of current LLMs force one to give another explanation for their performance that does not tie the ability to produce meaningful text to the capability of understanding meaning (Durt, Froese, and Fuchs 2023).

In assessing these phenomena, it is vital to distinguish between the methodologically rigorous science and engineering of LLMs, on the one hand, and philosophical and theological reflection *vis-à-vis* the more sensational popular, journalistic, and marketing discourse on technology, on the other hand. Some CEOs and engineers of companies who are developing those applications already market them as rising above the focused singular-task management of so-called 'narrow AI' and venturing into the realm of what has been termed 'artificial general intelligence' (AGI), that is: performing on a broad range of tasks: 'reasoning,' 'learning,' 'improving themselves,' and

3 In its contemporary rendering, the test is set up like this: A human person (P1) is chatting with another human person (P2) and an AI system. If P1, in this chat setup, cannot clearly differentiate who is P2 and who is the AI system, then that AI system has passed the test, and we have to assume that it is thinking and conscious.

showing forth human or even superhuman cognitive capabilities[4]. LLMs are thus framed either as the next step on our path to harness the promises of technology or, contrarily, as an existential threat to humanity[5].

But do current LLMs actually warrant such claims – both the hopes and the fears? Given the speed of development and the urgency with which a diverse array of views, interpretations, and predictions are being put forward, it seems vital to take a step back and ask fundamental questions about LLMs: How do these systems work? What are they currently capable of? Particularly: Do they really 'understand' the meaning of the texts they draw upon, process, and generate? What are their limitations? Are they comparable to human beings regarding their linguistic capabilities and language use? What does their performance tell us about human language, its use, and the human linguistic capacity more broadly? What would it mean to say that they are unlike human beings, but nevertheless capable of understanding and using language with meaningful insight?

Answers to such questions are relevant to our ethics, economics, and politics. At this stage, however, such answers are also necessarily speculative (except perhaps the one about the mechanics of LLMs) and require further conceptual work as well as empirical studies (Pavlick 2023). Such answers, furthermore, reach beyond engineering and practical ethics into philosophical and theological territories. They rest on basic assumptions about and models of the world, language, cognition, mind, and human beings, and, therefore, are open to interpretation and invite interdisciplinary scrutiny from a plurality of outlooks.

---

4 OpenAI, for example, the firm that introduced ChatGPT and GPT-4 (at the moment of writing this), advertises that their "research will eventually lead to artificial general intelligence, a system that can solve human-level problems" (see OpenAI. "Pioneering research on the path to AGI"). That such a view is more broadly held is evinced by an open letter – signed by thousands of leading AI researchers – which calls "on all AI labs" to "pause the training of AI systems more powerful than GPT-4." They ask: "*Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us" (see Future of Life Institute 2023, original emphasis).

5 One set of fears seems to be circling the idea that once such a system has reached 'human-level capabilities' on a broad scale, or: AGI, it will take on a life of its own, improve itself at a much faster rate than human development and thus become uncontrollable while at the same time coming up with its own goals and values which may not align with or be favourable for human beings. Variations on that theme have driven debates around 'AI safety,' 'AI alignment,' and the 'existential risks' (Bostrom 2013) of a coming 'superintelligence' (Bostrom 2017) or even a 'singularity' (Kurzweil 2005; Eden et al. 2012). For a critical review of such 'transhumanistic' predictions, see Dürr 2021 and Steinmann 2020.

This paper should, therefore, not be read as a definitive statement, but as a position paper, with a particular viewpoint. It aims to provoke further critical discussion on these issues – particularly the philosophical and theological presuppositions underlying the debate. In what follows, we want to compare two different – necessarily typecasted – approaches to making sense of cognition and linguistic meaning-making, which rest upon competing models of mind and human beings: the holistic-enaction paradigm (section 2.a) and the information-processing paradigm (section 2.b). We then introduce and assess two language generation models – one presently in use and one more speculative: firstly, the transformer architecture as used in current large language models, such as the GPT-series (section 3.a), and secondly, a model for 'autonomous machine intelligence' recently proposed by Yann LeCun, which involves not only language but a sensory-motor interaction with the world (section 3.b). We finally discuss the language capacity of these models from the perspectives of the two rivalling language paradigms (section 3.c) before we reach a conclusion (section 4).

2. What is the Meaning of Meaning? Rivalling Interpretations of Mind, Cognition, and Language

a) The Holistic-Enaction Paradigm

In what follows, we will not so much set forth a particular *theory* of human language but reflect on the conditions of linguistic meaning-making following a broadly 'enactivist' approach[6], which understands higher linguistic forms of cognition as grounded in the activities of living organisms (Varela, Thompson, and Rosch 1991; Thompson 2007; Stewart, Gapenne, and Di Paolo 2010; Gallagher 2011; Di Paolo, Buhrmann, and Barandiaran 2017; Fuchs 2018) as well as 'forms of life' from which linguistic expressions derive their particular meaning (Moyal-Sharrock 2022)[7]. This lies in tension with much of the history of the analytical philosophy of language, which, in its grappling with the truth conditions of language as a formal system, at least since Frege (Heck and May 2006), has paid almost no attention to

6 Enactivism is not a homogenous field of research – for a helpful overview of the 'varieties of enactivism' see Ward, Silverman, and Villalobos 2017.

7 In the remainder of this section, we will explain that such an approach views any kind of explicit (propositional) theorizing as grounded in more basic activities of the human bodily organism as well as a particular 'form of life,' which both are not explicitly propositional (even though they can be described in such terms) but rather practical pre-theoretical activity (Venturinha 2016, 115–16; Husserl 1976).

embodiment and non-verbal communication going along with propositional structures[8].

For enactivists, human linguistic capacity and thus their capacity for sense-making rests on being embodied (living and organic) and embedded in an environment with other such beings. In explaining why this is the case, enactivism works with core concepts like 'autonomy' and 'sense-making,' which help understand what 'meaning' means on different levels of organic existence. Let us briefly outline those concepts[9]:

In the enactivist view, what makes humans cognitive beings is that their bodies realise and sustain a certain kind of 'autonomy'[10] – they are "internally self-constructive" in a way that enables them to interact with and adapt to their environment actively (Thompson and Stapledon 2009, 24). Because 'autonomous' systems actively sustain themselves, they generate their own domains of meaning and value which corresponds to their efforts of self-conservation in precarious environments (Jonas 1966). Take, for example, motile bacteria which swim uphill on a food gradient of sugar (on this, see Thompson and Stapledon 2009, 24–25; Thompson 2007, 74–75, 157–58): At first, they tumble around randomly until they encounter an orientation that increases their input of sugar. At this point, they swim forward up the gradient toward the place with maximal sugar concentration. Because of the way these organisms' metabolism chemically realises its autonomy, more sugar is better for them. Sugar, to those bacteria, is significant and valuable. Those, however, are features related to these bacteria; they are not intrinsic to sugar molecules. Thus, living autonomous organisms transform their world into a "place of salience, meaning, and value" – the navigation of which is what enactivists refer to as "sense-making" (Thompson and

8 This corresponds with an influential interpretation of Ludwig Wittgenstein's philosophy of language as articulated in his assertion that all philosophy is "critique of language" (Wittgenstein 1961, 4.0031). The task of philosophy is then to clarify logical and semantic problems of language, ultimately to distinguish the sayable from the unsayable (Wittgenstein 1961, Preface, 4.114, 6.53).

9 Note that concepts like 'autonomy,' 'sense-making,' 'normativity' etc. are deployed here within the framework of (the philosophy or theory of) biological life, organisation and survival behaviour and must therefore not be confused with or prematurely identified with the meanings (or interpretations) of those terms, e. g., in philosophy or theology.

10 'Autonomy' or 'basic autonomy' is "the capacity of a system to manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control: (i) internal self-constructive processes and (ii) processes of exchange with the environment" (Ruiz-Mirazo and Moreno 2004, 240). This has also been explained in terms of their "operational closure" (Varela 1997) or, on the molecular level, as "autopoiesis" (Maturana and Varela 1980).

Stapledon 2009, 25)[11]. For autonomous sense-making organisms, their interactions with the world thus acquire 'normative status.' They modify their environment according to the inner norms of their self-sustaining activities along a spectrum of vitality (spanning from health, stress, fatigue, and sickness even to death). This proactive behaviour suggests intentionality (in the phenomenological sense) and purpose as inherent features of life and living organisms (Varela, Thompson, and Rosch 1991, 205–06; Thompson 2007, 26–27; Turner 2017; Noble and Noble 2023).

What does this mean for human beings and their linguistic capacity? The enactivist reading of the biology of living organisms submits that living beings actively sustain themselves via sense-making. The meaning of this behaviour is bound up with the significance and value the world has for *them*. With bacteria, this pertains to sugar levels – with 'higher' animals, like human beings, this can inform a host of more refined and complex forms of cognitive and even cooperative behaviours.

Meaning and sense-making are thus happening in a sphere that cannot really be localized in Euclidean space. In this sense, the enactivist approach cuts across internalist-externalist debates about the localisation of cognition: Because cognition is understood here as a continuous "relational process of sense-making that takes place between the system and its environment" (Thompson and Stapledon 2009, 25–29, here: 26; Maturana and Varela 1980), a closed feedback loop (Fuchs 2018). In contrast, the internalist-externalist debate rests on presuppositions entirely foreign to the enactivist approach. Enactivism builds upon (or is at least highly compatible with) how philosophers like Martin Heidegger and Maurice Merleau-Ponty conceive of the mind's relation to reality: Not as a relationship between a distinct subject and object, but rather as a way of "being-in-the-world" which amounts to the 'space' in which we become capable of coping with the world (Merleau-Ponty [1945] 2005; Heidegger [1926] 2006; Dreyfus 2016). Human cognition then does not happen 'in the body' alone, much less 'in the brain' in a localised sense, but in and between self, body, and environment (Fuchs 2018). However, it should be clear from what has been said so far that the living body plays a fundamental role in cognition: This is, firstly, because higher-level activities build upon its basic self-regulating metabolic activity, but also, secondly, because it is the body that places an individual in an environment as well as a world of other bodies with whom it inter-

11 Evan Thompson and Mog Stapledon define it as follows: "Sense-making is behaviour or conduct in relation to environmental significance and valence, which the organism itself enacts or brings forth on the basis of its autonomy" (Thompson and Stapledon 2009, 25).

acts in the process of sense-making. 'Sense-making' here manifests itself as multilayered, its meaning depending on the level of being (metabolism, organisational structure, psychology, interaction with the environment, or with other human beings), which is addressed.

Thus, interaction, cooperation, and even social forms of cognition come into view (De Jaegher and Di Paolo 2007, 485–507; De Jaegher, Di Paolo, and Gallagher 2010). Enactivism considers the kind of relationality that grounds enactive 'meaning' not only with regard to individuals but also in the social sphere of their co-existence. This corresponds, for example, with how language acquisition seems to happen in a shared space of "joint attention" or "communion" (Tomasello and Farrar 1986; Tomasello 2014; Mundy 2018). At this higher level of social cognition and interaction, language seems to play an important role. Language is a social phenomenon which unites persons, as living bodies, with others in a community. Possessing language *is* to be in a meaningful social space and in relation to others. Principally, this is the relation to any other linguistically gifted person; practically, it situates a human being in a certain family, culture, religious group, etc. The sum of those relations, in other words, the 'form of life' in which an individual is enmeshed, determines how they use language and what it means to them (Taylor 2016).

Such an enactivistically inspired paradigm is highly compatible with Ludwig Wittgenstein's late thought – particularly interpreted through the lens of *On Certainty* (Hutto 2013; Moyal-Sharrock 2013). In the enactivist reading of Wittgenstein, human linguistic capabilities, knowledge, and cognition (more broadly), all build upon and refine practical behaviour (Moyal-Sharrock 2022; Canfield 2007). This amounts to a grounding of cognition in social practices, which functions as a "corrective" to representational epistemologies (Venturhina 2016, 115; Moyal-Sharrock 2022, 185–98). Wittgenstein solves the problem of the foundation of knowledge with the introduction of 'basic beliefs,' which are not themselves based on further propositional beliefs (or else an infinite regress would ensue) but rather on "animal and unreflective ways of acting" (Moyal-Sharrock 2022, 185)[12]. Thus, Wittgenstein might be read as "perhaps the first enactivist" (Moyal-Sharrock 2013, 266) *avant la lettre*.

He helpfully compares using a language to playing a game (Wittgenstein 1997, 1969–1975): We use words then similarly to how we play tennis – often without theoretically 'knowing' the mechanics (Moyal Sharrock 2022, 41). Mastering a 'language-game' means mastering its 'grammar.' In the enactivist

12 Only if one formulates those beliefs can they seem to be (empirical) propositions.

reading of Wittgenstein, this is more than merely mastering the rules for the correct use of words (against Hacker 2012), but refers, in some cases, to embedded *ways of acting* which, conceptually elucidated, function as rules of grammar – that is: they underlie all thinking, underpin what we say and do and ultimately even delineate the bounds of sense for us[13]. The 'rules' or 'grammar' of a language-game "are simply expressions of the norms of sense that are socially generated and maintained; they grow out of, and with, our natural ways of acting and our socio-cultural practices" (Moyal-Sharrock 2022, 91)[14]. Wittgenstein makes this point by quoting Goethe's Faust: "In the beginning was the deed," which is preceded by the statement: "The origin and the primitive form of the language-game is a reaction; only from this can more complicated forms develop. Language – I want to say – is a refinement" (Wittgenstein 1993, 395). Thus, what connects "propositions" with "reality" and consequently conditions our understanding of "sense" goes beyond the mere handling of words (Wittgenstein 2005, § 42). It includes coping with the world and with others – because meaning exists *within* and results *from* the language use of a community (Wittgenstein 1997, 1969–1975)[15]. Wittgenstein expresses here an early form of a theory of speech acts (see later Austin 1962; Searle 1969) insofar as linguistic utterances within a language-game are always also actions (Wittgenstein 1997, § 27). In Wittgenstein's words: "the term 'language-game' is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life" (§ 23).

This embedding of 'meaning' in more encompassing contexts, which are nevertheless particularly situated in forms of life, is what we term a 'holistic-enactivist paradigm' of language. It reckons with *meanings of meaning* that unfold on multiple layers (biological, psychological, spiritual, etc.) and in multiple domains of reality (individual, social, etc.)[16]. Charles Taylor has a helpful way of differentiating meanings of meaning on the higher

13 Moyal-Sharrock, arguing here against Peter Hacker (Hacker 2012), speaks here of "hinge certainties" which sometimes function as rules of grammar (Moyal-Sharrock 2022).

14 It is not surprising, on this view, that language-games are found in a wide range of culturally varied forms. Precisely because meaningful speech involves being concretely situated, being anchored to a viewpoint, and living with limitations (being here and not somewhere else, saying this and not something else, saying it in this language and not in another etc.) it will take on different forms in different settings.

15 This connection between language-as-a-system and language-in-use is what already underlies Ferdinand de Saussures differentiation between *langue* (a system of signs and rules) and *parole* (the use of language in a particular situation). De Saussure observed that *parole* continuously changes *langue,* while at the same time *langue* structures *parole* (De Saussure [1916] 2011).

16 This corresponds with the above-mentioned, multilayered notion of 'sense-making.'

levels of animal and human life. He distinguishes 'human meanings' from more general 'life meanings' (Taylor 2016, 92–93): Life meanings refer to having the means to survive (food, shelter, and the like), which human beings share with other 'higher' forms of life. An interesting case of this, however, is the fact that human beings seek communion, intimacy, and love. In our phenomenological experience in the stream of life, the latter one lies somewhere between animalistic survival instincts and something beyond, which the term 'human meanings' tries to capture. Human persons can understand 'love' in a way that seems to transcend evolutionary instincts. They even linguistically articulate what it means to live up to the demands or come up with an ideal of love. Such phenomena are what drive, for example, ascetic religious practices like fasting, not sleeping, refraining from sexual activities, and the like. Such humanly meaningful behaviour at times seems to go against the survival-based 'sense-making' on the animalistic or even organic level. This is not to deny the value of analysing these phenomena from the methodological perspective – say, of evolutionary psychology (Buss 2018), enactivist approaches to social cooperation (De Jaegher and Di Paolo 2007, 485–507), or value theory (Fuchs 2019) – and much less to present our own theory. The point is rather to acknowledge that the meaning of things, for humans, can only ever be fully explained by considering their particular experience and interpretation thereof. Thus, while certain levels of meaning are objectively recognisable (Taylor's 'life meanings'), the language of 'human meanings' does not "translate into that of objectively identifiable states" without loss. Human meaning can never be grasped from the outside "without reference to the agent's self-understanding" (Taylor 2016, 92–93), which is inextricably linked with the "what-it-is-likeness" of their phenomenal experience (Nagel 1974, 2012). Both of those, from a holistic-enactivist perspective, are grounded in the morphology of their organic, living bodies.

In sum: From a holistic-enactivist perspective, 'Meaning' in the full human sense as it pertains to language can only be understood against a holistic background, which consists in the manifold ways in which a self is embodied and embedded and enacts the world[17]. Higher cognitive faculties

17 Hubert Dreyfus and Charles Taylor summarize much of what has been argued for here in terms of a "contact theory" with which they aim to retrieve realism: "A basic move which gives rise to this theory is a reembedding of thought and knowledge in the bodily and social-cultural contexts in which it takes place. The attempt is to articulate the framework or context within which our explicit depictions of reality make sense, and to show how this is inseparable from our activity as the kind of embodied, social, and cultural beings we are. The contact here is not achieved on the level of Ideas [sic!], but is rather something primordial, something we never escape. It is the contact of

can then be seen as grounded in more basic life processes and activities of the human body (as a living organism). They are inextricably tied up with 'forms of life' – referring here to the morphology and activity of living organisms, as well as the interpersonal and socio-cultural forms of living and interacting with others.

b) The Information-Processing Paradigm

An entirely different approach to cognition, language, and meaning-making is what we term an 'information-processing paradigm' (Floridi 2011, 2013, 2019; Munataka 2016; Berk 2018; Tegmark 2017)[18]. We begin with outlining certain assumptions that implicitly or explicitly underlie this line of reasoning. The aim of this sketch is to illuminate the differences between informational and holistic-enactivist approaches in how they would assess LLMs.

The information-processing paradigm assumes that minds receive 'inputs' from their environment, process them internally and produce corresponding 'outputs'; that these are essentially computational processes operating with clearly defined and explicit bits of information; that mental operations can be explained with reference to the physical operations of the underlying brain (the information-processing engine); and, finally, that brains, like computers, are 'syntactic engines'[19] referring to the world 'out there' through such 'inputs' and acting on it by way of such 'outputs.' Mark Tegmark provides a recent example of this when he describes minds, or "intelligent agents," as "entities that collect information about their environment from sensors and then process this information to decide how to act back on their environment" (Tegmark 2017, 25).

Building on the functionalist premise, "what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part" (Levin 2023), and thus the assumption of 'substrate

living, active beings, whose life form involves acting in and on a world which also acts on them. These beings are at grips with a world and each other; this original contact provides the sense-making context for all their knowledge constructions, which, however much they are based on mediating depictions, rely for their meaning on this primordial and indissoluble involvement in the surrounding reality" (Dreyfus and Taylor 2015, 18–19).

18 The term 'information' today is deployed in many areas of research in both the natural sciences and the humanities. We deploy it here, very broadly, as referring to the elements of a universal form of computation (= information processing).

19 On this, see Dreyfus and Taylor 2015, 15.

independence,' an information-processing account sees mind wherever meaningful behaviour (as in our case, the generation of meaningful text) goes along with observable structural isomorphies (Clark and Chalmers 1998). This corresponds with a computational theory of mind or a 'soft computationalism,' which holds that the right computational structure suffices for the possession of a mind ('computational sufficiency') and that this mind's cognitive processes are best understood within a general framework of computation (Chalmers 2011). Such views cut both ways and lend themselves to potentially considering current LLMs to be 'mindful' as well as viewing human minds as 'software-like.' This is precisely how Tegmark describes the human mind: It consists of "all the algorithms and knowledge that you use to process the information from your senses and divide what to do – everything from your ability to recognize your friends when you see them to your ability to walk, read, write, calculate, sing and tell jokes" (Tegmark 2017, 27).

The information-processing paradigm is compatible with both internalist and externalist accounts of cognition. The noteworthy focus on the brain[20] as a computation engine suggests proximity to internalism (Chalmers 2010; Churchland and Sejnowski 1992; Churchland 2013). Still, many proponents of an informational paradigm take an externalist stance, like, e.g., the "extended mind theory" (Clark and Chalmers 1998). However, this externalism still takes 'brain activity' as the paradigm for cognitive processes (Clark 2003, 2008). Only in a second step are they then looking for isomorphic factors outside the brain, which have a comparable function in the production of behaviour and are therefore regarded as being part of the cognitive process in terms of "information-processing mergers" (Clark 2003, 5; see Thompson and Stapledon 2009). The body is then viewed primarily in terms of its functional role[21]. The main feature of the body is to provide the "stable (though not permanently fixed) platform whose features and relations can be relied upon in the computation of certain information-processing solutions" (Clark 2008, 55–56). To explain cognition, an information-processing account refers to the body as simply "whatever plays these roles in a unified information-processing economy" (56–58). Extended mind theory holds the mind to be the overall balance of "a kind

20 Such 'cerebrocentrism' has come under attack (Fuchs 2021, 107–23) and risks falling prey to the "homunculus fallacy" (Kenny 1984, 125–36) or "mereological fallacy" of attributing capacities of the human being as a whole in its entirety to parts of it, like the brain (Bennet and Hacker 2022, 79–93).

21 Others believe that the body plays an altogether negligible role (Tegmark 2017), an idea which also drives technological futurism (Kurzweil 2005; Bostrom 2017).

of equal-partners dance between brain, body, and world" (56–57). This, however, should not be confused with an enactivist approach (see Gallagher 2018), in which the living body – including the brain – *is the basis* for this 'dance' because it actualizes the autonomy which lies at the basis of agency and sense-making (Thompson and Stapledon 2009, 28).

What, then, is language in an information-processing paradigm? The human linguistic capacity is taken to be the capacity of the brain to process information (Massaro 1975; Churchland 1980; Churchland 1989; Churchland and Sejnowski 1992; Clark 2010). Meaningful propositions are understood to 'represent' reality, and reality, in turn, is conceived of as states of affairs composed of atomised facts that can be grasped as data. The extracerebral world is 'represented' as information in the brain in a 'world model' which mirrors the outside reality, interacts with it, and is continuously refined through the adaptive adjustment of inputs and outputs. Language, in this view, is the ability to encode, process, and decode aspects of reality – both mental and material – as bits of information. Language allows for the communication between the brain's 'inner' world model and the reality 'out there,' allowing for more effective ways to manipulate that world.

3. Assessing Artificial Language Generation

Having outlined two rivalling approaches to explain the human linguistic capacity, we now turn to current AI language generation on two levels. First, we briefly introduce and assess the language generated by today's powerful LLMs (section 3.a). We then take a closer look at the potential language of a recent proposal for 'autonomous machine intelligence' by Yann LeCun – considered to be one of the fathers of AI because of his groundbreaking work on deep learning[22] – which takes AI decidedly in the direction of human-like or even superhuman agency (section 3.b). Based on these two examples, we show how today's and potential future artificial languages can be interpreted from either a holistic-enactivist or an informational perspective. Finally, we take a holistic-enactivist stance and argue that LLMs and even LeCun's future AI do not represent a significant step toward a human-like language understanding in machines.

22 For this work he has won, together with Yoshua Bengio and Geoffrey Hinton, the ACM A. M. Turing Award (https://www.acm.org/media-center/2019/march/turing-award-2019).

a) Large Language Models and the Generating of Meaningful Text

Today's LLMs (such as the GPT-series by OpenAI) are so-called 'autoregressive transformer models[23].' An *autoregressive* model predicts the value of a variable based on the joint probability between past values (inputs) and possible next values (outputs). A joint probability might, for example, be the probability for the word "blue" given the preceding words "She paints the wall …" Instead of just maximising the joint probability and aiming for the most probable next value, these models usually sample from the joint probability density, yielding different, but still reasonable results to equal past values. This adds a random component such that LLMs always generate new answers (output) to equal text prompts (input). Much of the current success of LLMs is due to the introduction of the *transformer* model architecture, which makes efficient use of a mechanism called 'attention' (Vaswani et al. 2017; Devlin et al. 2019). In contrast to traditional neural networks for sequences of data (e. g., Hochreiter and Schmidhuber 1997; Hochreiter et al. 2001), transformers can model relations between values independently of the distance between them and allow for efficient parallelization, such that training on a large amount of data is feasible. In other words, the great innovation of transformers lies in their ability to connect and relate words and parts of words to each other, even if there is a great distance between them in a given text, and their probabilistic output prediction is based on large and incredibly representative statistical models.

It is important to note that in reality, the values of LLMs do not represent words but so-called sub-word 'tokens,' which refer to statistically relevant parts of a text, i. e., sequences of letters (including punctuation and spaces) that appear often enough. In order to subject a sequence of letters to a mathematical model, it must be numerically represented. This is done by identifying every token with a specific point in data space, e. g., "paint" with a point at coordinates (1.23, 4.23, 0.11) in three-dimensional data space. Note that the connections between tokens and points in space are not fixed in an arbitrary manner but learned during training. Furthermore, this data space is usually high-dimensional, e. g., 1024-dimensional, so we cannot intuitively grasp where a point lies or how it is being transformed during processing. Although input and output can be meaningfully interpreted, what happens inside a transformer remains mostly hidden and is learnt based on the

23 For a timely, more detailed account, we suggest Russell and Norvig 2021 and Murphy 2023.

input-output sample pairs provided during training[24]. The result is an astonishingly good 'next token predictor,' which generates, in many cases, convincing textual output.

The capacity of LLMs to generate text that is broadly perceived as meaningful is a major accomplishment in computer science and must be given credit. LLMs have even mastered so-called 'Winograd schemas' that were proposed in 2011 to measure the depth of understanding of language models as a qualification for the Turing test (Kocijan et al. 2023; Levesque 2011). Consider the sentence: "The car overtook the school bus because it was very fast." The fact that "it" refers to the car does not follow from syntax alone but from the words themselves. Thus, simple rule-based models have long struggled to proceed meaningfully from such prompts. But due to the attention mechanism, LLMs can now correctly connect "it" to "car" and thus output reasonable text, such as "So it eventually left the school bus far behind."

Currently, there is no consensus on whether today's LLMs represent a significant step toward an intrinsic understanding by something like AGI[25]. A research unit at Microsoft claims to already see "sparks" of it in OpenAI's GPT-4: "Given the breadth and depth of GPT-4's capabilities," they write, "we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system" (Bubeck et al. 2023). Blaise Agüera y Arcas, from Google Research, likewise argues that with LLMs "statistics do amount to understanding" and that they, therefore, "represent a major advance in artificial intelligence and, in particular, toward the goal of human-like artificial general intelligence" (Agüera y Arcas 2022). Despite such statements (not least from researchers working for companies who market their technology), there is a broad consensus in AI research, the philosophy of mind, and cognitive science that current LLMs are not sentient in a way that allows for human-like language understanding. The operational structure of LLMs suggests that its language generation is not achieved by LLMs somehow rising above syntax and getting at semantics, developing a true understanding of words (such as "car" and "school bus" in the above example). Instead, LLMs remain syntactic machines that solve Winograd schemas by expanding syntax from basic rules to very complex rules, enabling them to consider statistical correlations in large amounts

24 Although, there are efforts to visualize data mappings and dependencies in transformers that give a good and accessible intuition for how transformers work (Braşoveanu and Andonie 2020; Vig 2019; Van Aken et al. 2020).

25 See, e. g., Strickland and Zorpette 2023. Note that this article must be taken with a grain of salt, since the respective experts did not sign on their given opinion.

of data (Bender et al. 2021). Thus, current LLMs essentially work with the patterns and meanings sedimented in the (humanly authored) texts upon which those systems are trained and recombine and imitate human language rather than grasping its meaning intrinsically[26]. Yann LeCun, who is a leading figure in AI research, finds a lack of contact with the real world in current AI systems (LeCun 2022), leading to what Eric Larsson has called a 'simplified world' that does not imply any sign of understanding at all (Larson 2022). Together with many other AI researchers, LeCun also recognizes that current AI systems, based on artificial neural networks, are prone to instability, which is why there always exist examples (engineered as so-called "adversarial examples," see Goodfellow, Shlens, and Szegedy 2014) which reveal that there is no real-world understanding of the basic meaning of words and texts (Marcus and Davis 2019). Because this problem is a problem inherent to current AI model architectures, LeCun argues that it cannot be solved by upscaling toward larger models and larger training data sets. It is thus hard – from a holistic-enactivist but even from an internalist informational perspective – to argue for human-like language use in current LLMs: They lack structural isomorphies with the human mind, and even their meaningful behaviour rests on unstable grounds.

Interestingly, LeCun has recently proposed a new model for 'autonomous machine intelligence' that aims to overcome those limitations and toward more human-like understanding. He presents his vision for this in a position paper titled "A Path Towards Autonomous Machine Intelligence" (LeCun 2022)[27].

b) Robotic Embodiment and a Path Toward Autonomous Machine Intelligence

LeCun begins with the observation that the learning, understanding, and generalizing capabilities of humans and animals still far exceed those of today's AI systems. A human, for example, is much more reliable in real-world tasks, like driving, and needs very few trials to learn them, while an

26 This is one of the most striking things the success of LLMs demonstrates, namely: How much of the real-world context is contained in human language the patterns of meaningful text, and that with enough such data, we are capable of engineering systems that can generate not only the syntax of language, but also such meaningful patterns of language use (Durt, Froese, and Fuchs 2023). Their grounding in humanly authored text, however, is the basis for current discussions about copyright safety for generative AI (Sag 2023).

27 For a more popular and accessible rendering of his approach see Browning and Lecun 2022.

AI system has to be trained on massive amounts of data (LeCun 2022, 2). Drawing from psychology (Craik [1943] 1952; Bryson and Ho 1969) and cognitive science (Lake et al. 2017; Orhan, Gupta, and Lake 2020), LeCun traces this difference back to variations in what he calls the learned "world models", i. e. "internal models of how the world works" (LeCun 2022, 3). In his view, humans acquire enormous amounts of background knowledge that feed into their world models, leading them to be able to predict and reason more efficiently in new situations. Thus, he regards world models to be the basis of what is often called 'common-sense' and, therefore, an important steppingstone on the path toward autonomous machine intelligence[28].

This, argues LeCun, is precisely something that LLMs lack (LeCun 2022, 45): Even though much background knowledge can be "extracted" from the written texts on which these systems are trained, most of what makes up human common-sense is not represented in any text but rather "results" from interaction with the physical world. Thus, for LeCun, it is because "LLMs have no direct experience with an underlying reality," that "the type of common-sense knowledge they exhibit is very shallow and can be disconnected from reality" (45).

Based on his assumptions about world model common-sense, LeCun proposes a new direction in AI. He argues that new model architectures that better resemble brain function must be innovated if we want to achieve autonomous intelligent agents. His proposition for this involves six modules: A *perception module* estimates the current state of the world with the use of sensors in a hierarchical fashion. A *world model module* has two tasks: firstly, it estimates missing information about the world from the perception module; secondly, it predicts future states of the world. A *cost module* measures current 'discomfort' in the agent and predicts future cost states. A *short-term memory module* stores and updates information. An *actor module* computes proposals for actions to the world module and feeds it into the cost module, thereby being able to minimize future 'discomfort' by optimizing actions sent to the effectors. A *configurator module,* finally, is connected to all other modules and configures them concerning a particular task at hand.

In contrast to current AI models (like LLMs), LeCun's model is not limited to predicting language using sequences of tokens but operates with

28 'Common-sense' refers to the practical judgments concerning everyday matters that require a basic understanding shared by almost all people. This has been described as one of the major challenges to AI-reasoning (e. g., Levesque 2017; Marcus and Davis 2019).

a very general cost function (the goal of reducing 'discomfort'), which allows for 'learning' mechanisms, the setting of various sub-goals and thus increasing 'autonomy[29].' Furthermore, it is connected to and interacts with the material world through access sensors and effectors.

LeCun's approach has already been positively taken up and thus warrants closer inspection (Matsuo et al. 2022). It is a particularly interesting position paper because it situates technical propositions within theoretical frameworks of mind, human understanding, and 'common-sense' and thus invites philosophical discussion. LeCun – like others (Bisk et al. 2020; Bender and Koller 2020; Marcus, Leivada, and Elliot 2023) – takes up some enactivist considerations: namely that common-sense, understanding, and the meaning of language are grounded in embodied experience in the material world. It is unclear, however, what 'embodiment' in such accounts actually means and what role the bodies' substrate plays for the experiential grounding of cognition.

It seems that LeCun remains broadly within an informational paradigm. His line of argument suggests that the reason why large areas of human common-sense are not "represented" in written text is just that they happen not to have been written up, but potentially could be – rather than supposing that they really are grounded in what could be termed 'non-' or 'pre-propositional' enaction, navigation and coping (Dreyfus 2012, 2016). In a holistic-enactivist perspective, explicit propositional language is always embedded in such a wider background, refers to it, and draws its meaning from it, but it, in principle, never exhausts it. The pressing question with artificial language models is just how much of this background *can* be represented digitally. LeCun seems to take the view that all aspects of common-sense that lie, so to speak, *in* the bodily actions themselves can, in principle, be included in a representational model – which implies that those aspects can be encoded informationally, or at least: that such encoding is exhaustive enough for the practical purposes of creating an autonomous machine intelligence.

This is corroborated by his focus on a 'world model' *internal* to the system, which (through a perception module) is continually updated through sensory inputs and (through an actor module) continually computes sequences of actions and outputs them to effectors. This is a representational model insofar as the relevant features of the interaction with the external reality

29 It is vital to note here that terms like 'learning' and 'autonomy' have particular technical meanings that do not directly match the meaning of these terms when employed in enactivism or philosophy. In the technical context here, 'autonomy' basically means 'minimal human involvement in defining and reaching goals.'

(i. e., the material world) are being represented in the system's world model. Thus, the processing ultimately happens in the representationalist world model, which informationally mirrors the world 'out there.' Similarly, but in a more explicit fashion, Andy Clark argues that brains are best described as situated "prediction machines" (Clark 2013, 181). According to Clark, there is much reason to believe that the main function of the brain is error correction. Instead of using 'discomfort' to measure the error, Clark proposes "surprisal," i. e., the distance between what the brain predicts and what it perceives (186). The brain has input and output channels, but "all that it 'knows', in any direct sense, are the ways its own states (e. g., spike trains) flow and alter" (183). As with LeCun's proposal, the perceived world is an "internal model of the source of the signals: the world hidden behind the veil of perception" (184). Thus, minimising 'surprisal' amounts to "improving the world model so as to reduce prediction errors, hence reducing surprisal" (186). In a similar fashion, the brain controls motor action on corresponding output channels in such a way that minimises the difference between the imagined future state and the actual state by finding the set of intermediate states that lead from the actual state to the future state (186). In sum, "perception, cognition, and action … work closely together to minimise sensory prediction errors by selectively sampling, and actively sculpting, the stimulus array" (186).

For LeCun, as for Clark, the function of the body seems to be merely that of transmitting (encoding and decoding) inputs and outputs between the calculating system and the surrounding world – its biological organization is not seen to have any particular contribution to the process. This is broadly in line with the "mechanistic embodiment" (Sharkey and Ziemke 2001, 253–54) of what has been termed "robotic functionalism" (252) and how informational-accounts think of the role of the body in cognition (Clark 2008; Tegmark 2017), which consider it "merely a contingent (and increasingly negotiable) fact about human embodiment that the body is … metabolic" (Clark 2008, 56). Whatever is "special" about the body should be understood "through the familiar lens of our best information-processing models of mind and cognition" (58). LeCun is also compatible with more cautious outlooks, which aim to integrate bodily 'grounding' by modelling it with sufficient structural similarity within the representational model (Pavlick 2023; Piantadosi and Hill 2022). Following the close connection between LeCun's model proposal and Clark's informational account of the extended human mind, it can be said that, from an informational perspective, LeCun's vision, if implemented, would represent a significant step toward engineering human-like language. In the remainder of this chapter, we, however, argue

from a holistic-enactivist perspective that the informational paradigm overlooks enactive conditions for words and language to bear any significance.

b) Meaning, Significance, and Conviviality

LeCun's approach to overcoming the limitations of current LLMs brings to the fore a crucial feature of the informational paradigm: *It does not account for the role of vital embodiment in the making and understanding of meaning.* This becomes clear if one reckons with the role of *feeling and emotion* in cognition. LeCun speculates that "emotions" are the "result of brain structures" that play a similar role to the "cost module" of his proposed architecture (LeCun 2022, 44). Emotions are taken here to be the "product of an intrinsic cost," or the "anticipation of outcomes" calculated by this module. Therefore, LeCun writes, an "autonomous intelligent agent" possessing such a module "will inevitably possess the equivalent of emotions" (44–45). Such a rendering of what comprises emotion is compatible with stances that understand cognition as essentially a substrate-independent and affectless form of information-processing or problem-solving (Adams and Aizawa 2008; Clark 2007; sections 2.b and 3.b).

From an enactivist perspective, feeling is not something that can be modelled on a computer or in a robot because it is a vital part of ongoing "vital regulatory processes" involving the whole organism and its environment (Fuchs 2018). As we have already seen above, enactivism understands cognition and emotion as integrated at the biological, psychological, and phenomenological levels (Colombetti and Thompson 2005; Fuchs 2022). Cognition is grounded in the metabolism, autonomous organization, and sense-making activity of living beings for whom the material world acquires significance because they are navigating and coping with it. Here, cognition and emotion converge: "What has salience and value also has valence: it attracts or repels, elicits approach or avoidance. Such action tendencies in relation to value are the basis of emotion" (Thompson and Stapledon 2009, 26). This holds on all levels of description, from the bodily-affective feeling of life and primary self-awareness all the way up to higher mental functions, conscious experience, and self-experience (Fuchs 2022; Jonas 1966; Thompson 2007). Thus, what something means to *me* – in a holistic sense – depends on my vital embodiment and the higher emotional and cognitive functions resting on it. Embodiment, furthermore, always also places me in a specific context and provides me with a specific affectively coloured perspective. It is these concrete restraints which give some things

significance and valence and make others irrelevant and unattractive for *me* – on a biological, psychological, and phenomenological level.

All of these factors comprise the form of life in which meaning gains its meanings. For enactivists, this holds also on the socio-cultural level: Understanding why something has significance for another, in the fullest sense, then requires someone to share with them their form of life – on a biological as well as on a social level – which generates this significance. The form of life is what gives something (like a text) its particular meaning. In this view, understanding words in a way that allows for meaningful communication – on both ends – would require such a sharing of forms of life, which has been called "conviviality" (Zahavi 2015; Fuchs 2022). A good example of this is the term 'hunger' – the phenomenon and thus the meaning of the term can only be understood if 'hunger' is felt, and it can only be felt by an embodied living being subjected to metabolism, to constant vital regulatory processes that involve the whole biological organism and its environment (Damasio 2010; Fuchs 2018)[30]. Thus, however sophisticated and impressive an implementation of LeCun's vision for the future of AI would look like and behave, there is no reason to assume that if it outputs something like "I am hungry," it has some grasp on what humans mean when they say these words. If a system does not share our form of life – socially and biologically – it seems that it cannot understand us in a holistic sense (section 4).

In sum: Bodiliness, aliveness, and thus a vitally embodied wholeness are the conditions of possibility for the concrete understanding of meaning (particularly human meanings) and thus constitute a qualitative difference between human beings, on the one hand, and LLMs as well as sophisticated robots on the other hand.

4. Conclusion and Outlook

We have outlined how differently LLMs and LeCun's approach to autonomous AI are interpreted, depending on if one takes a broadly holistic-enactivist or an information-processing approach to interpret cognition, emotion, language, and meaning. As should have become apparent, we are sympathetic to an enactivist reading, in which only vitally embodied beings – animals, persons – generate, understand, and communicate mean-

30 The only conceivable way of integrating this yet again in an informational account seems to be to adopt a fully informational metaphysics, which considers the world in which we live and share our lives as one big system of information (Wiener 1948; Wheeler 1990; Capurro 2017; Floridi 2011).

ing in the strict sense. Only such beings properly 'have' intentions, aims to achieve and problems to solve in the process of reaching their goals[31]. The narrow kind of information-processing and problem-solving we delegate to AI systems in the sphere of human meaning presupposes the broader emotive cognition of sense-making and adaptive sense-regulation as well as a particular form of life, that we do not share with those systems (against Ziemke 2016, 2020; Hellström and Bensch 2018; Cappelen and Dever 2021).

The enactivist criterion of vital embodiment as a basis for cognition and emotion suggests hesitation toward attributing those properties to LLMs or other AI systems like the one proposed by LeCun. Because they are not grounded in life, it seems that nothing actually can matter to *them,* and therefore, they would not produce value or intentionality (Zlatev 2003; Fuchs 2022)[32]. This, however, is a necessary condition for meaning on the holistic-enactivist account we have sketched above. In contrast to AI systems, humans can differentiate between meaningless noise and meaningful signal (Tallis 2020, 240; 2004), which in the technical sense of 'information' in communication theory can be identical because information here is entirely devoid of meaning (Shannon and Weaver [1949] 1998, 8). So, there is no magical difference between patterns in which one cannot discern a face and patterns in which one can, but there is a categorical difference between entities to whom that difference means something and to such for whom it does not (Scruton 2019, 44). Living, embodied beings which existentially negotiate meaning in specific situations and environments and do so in communicative relation to others, really do discern signals, significance, and meaning in all the noise.

At the highest level of sense-making – including and building upon all the lower levels – the production of meaning, individually and collectively, seems to be a particularly human feature. Language plays a mediating role here as it structures meaning in the realm of human experience but also is a material process, "something that bodies do" (Williams 2014, x). Perhaps language is appropriately characterised as the way in which human persons mediate these two dimensions – that they are capable of doing this in the first place, is one of the most interesting features of our universe (Williams 2014). Such mediation is possible because human beings are both holistic

31 This presupposes that 'having goals' can be described in more specific terms than merely exhibiting goal-oriented behaviour and furthermore considers how a certain 'goal' emerges as such (against Tegmark 2017).

32 Thomas Fuchs elaborates why, in this view, what in engineering is termed 'artificial life' is not sufficient in providing the necessary vitality of embodied life which grounds cognition and the generation of meaning (Fuchs 2022).

beings and embodied; their bodies are simultaneously material things and saturated with meaning because they are bodies of *living persons*. Having this in mind opens up the perspective for the meaningfulness of extra-linguistic (in the strict sense) bodily activities like gesture in which language is grounded but for which an information-processing account has no real place (Clough and Duff 2020).

The fact that LLMs are not alive, together with consideration of their operational structure, suggests that they do not intrinsically 'understand' the meaning of the texts they generate, but rather essentially work with the patterns, meanings, and real-world embeddings condensed in human text. Playing with Frege's terminology, one might say that the outputs LLMs generate make 'sense' (*Sinn*) but have no (or at least only indirect) 'meaning' (*Bedeutung*) because they do not live in or refer to the extra-linguistic world of living bodies (Frege 1892; against Piantadosi and Hill 2022). The intelligent behaviour of such systems is best conceived of as an extension of the intelligence of designers, programmers, users, and authors of training data – and what is sedimented of this in their artefacts. In this sense, LLMs only ever solve *our* problems (or generate them), and only if we prompt them to do so.

This, however, only highlights the need for further research – empirical, conceptual, theoretical – into why LLMs are capable of producing (to some degree) meaningful text, without understanding meaning. Furthermore: what this fact tells us about language, human beings, and AI. We suggest that in answering such questions, proponents of different (philosophical) approaches – which we have outlined as informational and holistic-enactivist paradigms – should engage in (a) constructive and critical dialogue, including other disciplines, particularly analysing the nature and role of life, information, and the metaphysics of being more generally. Such foundational and interdisciplinary work should (b) subsequently be made accessible to wider, non-technical, audiences as well as journalistic outlets covering these topics. Both efforts, we hope, will contribute to establish a conversation which includes all those affected by the digital transformation, and which would be in the service of humanity.

References

Adams, Fred, and Kenneth Aizawa. 2009. "Why the Mind is Still in the Head." In *The Cambridge Handbook of Situated Cognition,* edited by Murat Aydede and P. Robbins, 78–95. Cambridge: Cambridge University Press.

Agüera y Arcas, Blaise. 2022. "Do Large Language Models Understand Us?" *Daedalus* 151:183–97. Accessed August 28, 2023. https://doi.org/10.1162/daed_a_01909.

Austin, John. 1962. *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*. Oxford: Clarendon Press.

Bender, Emily, and Alexander Koller. 2020. "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–98. Online: ACL. Accessed August 28, 2023. https://aclanthology.org/2020.acl-main.463.pdf.

Bender, Emily, Angelina McMillan-Major, Timnit Gebru, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–23. New York, NY: ACM. Accessed August 28, 2023. https://doi.org/10.1145/3442188.3445922.

Bennett, Maxwell R., and Peter M. S. Hacker. 2022. *Philosophical Foundations of Neuroscience: Second Edition.* Hoboken: Wiley.

Berk, Laura. 2018. *Development Through the Lifespan.* 7th ed. Thousand Oaks: Sage Publications.

Bisk, Yonatan, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. 2020. "Experience Grounds Language." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8718–35. Online: ACL. Accessed August 28, 2023. https://aclanthology.org/2020.emnlp-main.703.pdf.

Bostrom, Nick. 2013. "Existential Risk Prevention as Global Priority." *Global Policy* 4:15–31.

–. 2017. *Superintelligence: Paths, Dangers, Strategies*. Reprinted with corrections. Oxford: Oxford University Press.

Braşoveanu, Adrian M. P., and Răzvan Andonie. 2020. "Visualizing Transformers for NLP: A Brief Survey." In: *Proceedings of the 2020 24th International Conference Information Visualisation (IV)*, Melbourne, Australia, September 7–11, 2020, 270–79. Accessed August 28, 2023. doi:10.1109/IV51561.2020.00051.

Browning, Jacob, and Yann Lecun. 2022. "AI and the Limits of Language." *Noema*, August 23, 2022. Accessed August 28, 2023. https://www.noemamag.com/ai-and-the-limits-of-language/.

Bryson, Arthur, and Yu-Chi Ho. 1969. *Applied Optimal Control: Optimization, Estimation, and Control*. Waltham, MA: Blaisdell.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, John A. Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuan-Fang Li, Scott M. Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. "Sparks of Artificial General Intelligence: Early experiments with GPT-4." *ArXiv abs/2303.12712*. Accessed August 28, 2023. https://arxiv.org/pdf/2303.12712.pdf.

Buss, David. 2018. "The Evolution of Love in Humans." In *The New Psychology of Love*, edited by Robert Sternberg and Karin Sternberg. 2nd ed. 42–63. Cambridge: Cambridge University Press.

Canfield, John V. 2007. *Becoming Human: The Development of Language, Self and Self-Consciousness*. Basingstoke: Macmillan.

Cappelen, Herman, and Josh Dever. 2021. *Making AI Intelligible: Philosophical Foundations*. Oxford: Oxford University Press.
Capurro, Rafael. 2017. *Homo Digitalis: Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik*. Wiesbaden: Springer.
Chalmers, David. 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17:7–65.
–. 2011. "A Computational Foundation for the Study of Cognition." *The Journal of Cognitive Science* 12:323–57.
Churchland, Patricia Smith, and Terrence Sejnowski. 1992. *The Computational Brain.* Cambridge, MA: MIT.
Churchland, Patricia Smith. 1980. "Language, Thought, and Information Processing." *Nous* 14:147–70.
–. 2013. *Touching a Nerve: The Self as Brain*. New York: Norton.
Churchland, Paul M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: MIT.
Clark Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58:7–19.
Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence.* Oxford: Oxford University Press.
–. 2007. "Curing Cognitive Hiccups: A Defense of the Extended Mind." *The Journal of Philosophy* 104:163–92.
–. 2008. "Pressing the Flesh: A Tension in the Study of the Embodied, Embedded Mind?" *Philosophy and Phenomenological Research* 76:37–59.
–. 2010. "Magic Words: How Language Augments Human Computation." In *Thought and Language*, edited by Jill Boucher and Peter Carruthers, 162–83. Cambridge: Cambridge University Press.
–. 2013. "Whatever Next? Predictive Brains, Situated Agents, and The Future of Cognitive Science." *Behavioral and Brain Sciences* 36:181–253.
Clough, Sharice, and Melissa C. Duff. 2020. "The Role of Gesture in Communication and Cognition: Implications for Understanding and Treating Neurogenic Communication Disorders." *Frontiers in Human Neuroscience* 14:1–22.
Colombetti, Giovanna, and Evan Thompson. 2005. "Enacting emotional interpretations with feeling." *Behavioral and Brain Sciences* 28:200–01.
Craik, Kenneth. [1943] 1952. *The Nature of Explanation*. Cambridge: Cambridge University Press.
Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven: Yale University Press.
Damasio, Antonio. 2010. *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon Books/Random House.
De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. "Participatory Sense-Making: An Enactive Approach to Social Cognition." *Phenomenology and the Cognitive Sciences* 6:485–507.
De Jaegher, Hanne, Ezequiel Di Paolo, and Shaun Gallagher. 2010. "Can Social Interaction Constitute Social Cognition?" *Trends in Cognitive Sciences* 14:441–47. Accessed August 29, 2023. https://doi.org/10.1016/j.tics.2010.06.009.
De Saussure, Ferdinand. [1916] 2011. *Course in General Linguistics*, translated by Wade Baskin, edited by Perry Meisel and Haun Saussy. New York: Columbia University Press.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis: ACL. Accessed 29, 2023. https://arxiv.org/abs/1810.04805.

Di Paolo Ezequiel, Thomas Buhrmann, Xabier Barandiaran. 2017. *Sensorimotor Life: An Enactive Proposal*. Oxford: Oxford University Press.

Dreyfus, Hubert, and Charles Taylor. 2015. *Retrieving Realism*. Cambridge, MA: Harvard University Press.

Dreyfus, Hubert. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason.* Cambridge, MA: MIT.

–. 2012. "A History of First Step Fallacies." *Minds & Machines* 22:87–99.

–. 2016. *Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action.* Oxford: Oxford University Press.

Dürr, Oliver. 2021. *Homo Novus: Vollendlichkeit in Zeitalter des Transhumanismus.* Münster: Aschendorff.

Durt, Christoph, Tom Froese, and Thomas Fuchs. 2023. "Against AI Understanding and Senence: Large Language Models, Meaning, and the Patterns of Human Language Use." Reprint. Accessed August 29, 2023. http://philsci-archive.pitt.edu/21983/.

Eden, Amnon H., Eric Steinhart, David Pearce, and James H. Moor. 2012. "Singularity Hypotheses: An Overview". In *Singularity Hypotheses: A Scientific and Philosophical Assessment*, edited by Amnon H. Eden, James H. Moor, Johnny H. Søraker, and Eric Steinhart, 1–12. Berlin; Heidelberg: Springer.

Floridi, Luciano. 2011. *The Philosophy of Information.* Oxford: Oxford University Press.

–. 2013. *The Ethics of Information.* Oxford: Oxford University Press.

–. 2019. *The Logic of Information.* Oxford: Oxford University Press.

Frege, Gottlob. 1892. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik* 100:25–50.

Fuchs, Thomas. 2018. *Ecology of the Brain. The Phenomenology and Biology of the Embodied Mind*. Oxford: Oxford University Press.

–. 2019. "Values as Relational Phenomena – A Sketch of an Enactive Theory of Value." In *Perceiving Truth and Value: Interdisciplinary Discussions on Perception as the Foundation of Ethics*, edited by Markus Mühling, David Gilland, and Yvonne Fördster, 23–42. Göttingen: Vandenhoeck & Ruprecht.

–. 2021. *In Defense of the Human Being: Foundational Questions of an Embodied Anthropology*. Oxford: Oxford University Press.

–. 2022. "Understanding Sophia? On Human Interaction with Artificial Agents." *Phenomenology and the Cognitive Sciences* :1–22. Accessed August 29, 2023. https://doi.org/10.1007/s11097-022-09848-0.

Future of Life Institute. 2023. "Pause Giant AI Experiments: An Open Letter." Published March 22, 2023. Accessed August 9, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Gallagher, Shaun. 2011. "Interpretations of Embodied Cognition." In *The Implications of Embodiment: Cognition and Communication*, edited by Wolfgang Tschacher and Claudia Bergomi, 59–70, Exeter: Imprint Academic.

Gallagher, Shaun. 2018. "The Extended Mind: State of the Question." *Southern Journal of Philosophy* 56:421–47.

Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. 2014. "Explaining and Harnessing Adversarial Examples." *arXiv:1412.6572*. Accessed August 29, 2023. https://arxiv.org/abs/1412.6572.

Hacker, Peter. 2012. "Wittgenstein on Grammar, Theses and Dogmatism." *Philosophical Investigations* 35:1–17.

Heck, Richard, and Robert May. 2006. "Frege's Contribution to Philosophy of Language." In *The Oxford Handbook of Philosophy of Language*, edited by Barry Smith and Ernest Lepore, 3–39. Oxford: Oxford University Press.

Heidegger, Martin. [1926] 2006. *Sein und Zeit*. 19th ed. Tübingen: Niemeyer.

Hellström, Thomas, and Suna Bensch. 2018. "Understandable Robots – What, Why, and How." *Paladyn, Journal of Behavioral Robotics* 9:110–23.

Hochreiter, Josef, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9:1735–80. Accessed August 29, 2023. doi:10.1162/neco.1997.9.8.1735.

Hochreiter, Sepp, Yoshua Bengio, Paolo Frasconi, and Jürgen Schmidhuber. 2001. "Gradient Flow in recurrent Nets: the Difficulty of Learning Long-Term Dependencies." Accessed August 30, 2023. https://ml.jku.at/publications/older/ch7.pdf.

Hoff, Johannes. 2021. *Verteidigung des Heiligen: Anthropologie der digitalen Transformation*. Freiburg i. Br.: Herder.

Husserl, Edmund. 1976. *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, 2nd ed. Vol. 6 of *Husserliana Edmund Husserl – Gesammelte Werke*, edited by Walter Biemel. Dordrecht: Springer.

Hutto, Daniel. 2013. "Enactivism, from a Wittgensteinian Point of View." *American Philosophical Quarterly* 50:281–302.

Jonas, Hans. 1966. *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston, IL: Northwestern University Press.

Kenny, Anthony. 1984. *The Legacy of Wittgenstein*. Oxford; New York: Blackwell.

Kocijan, Vid, Ernest Davis, Thomas Lukasiewicz, Gary Marcus, and Leora Morgenstern. 2023. "The Defeat of the Winograd Schema Challenge." *Artificial Intelligence*. Rreprint. Accessed August 29, 2023. https://arxiv.org/pdf/2201.02387.pdf.

Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Penguin.

Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017. "Building Machines That Learn and Think Like People." *Behavioral and Brain Sciences* 40. Accessed August 29, 2023. https://arxiv.org/abs/1604.00289.

Larson, Erik. 2022. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Harvard: Harvard University Press.

LeCun, Yann. 2022. "A Path Towards Autonomous Machine Intelligence." *OpenReview*. Accessed August 29, 2023. https://openreview.net/pdf?id=BZ5a1r-kVsf.

Levesque, Hector. 2011. "The Winograd Schema Challenge". In *Logical Formalisations of Commonsense Reasoning*, Papers from the 2011 AAAI Spring Symposium, Vol. 6, 63–68. Palo Alto, CA.

–. 2017. *Common Sense, the Turing Test, and the Quest for Real AI: Reflections on Natural and Artificial Intelligence*. Cambridge, MA; London: MIT.

Levin, Jane. 2023. "Functionalism." In *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), edited by Edward Zalta and Uri Nodelman. Accessed August 29, 2023. https://plato.stanford.edu/archives/sum2023/entries/functionalism/.

Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon.

Marcus, Gary, Evelina Leivada, and Elliot Murphy. 2023. *A Sentence is Worth a Thousand Pictures: Can Large Language Models Understand Human Language?* Unpublished position paper. Accessed August 29, 2023. https://arxiv.org/pdf/2308.00109.pdf.

Massaro, Dominic W., ed. 1975. *Understanding Language: An Information-Processing Analysis of Speech Perception, Reading, and Psycholinguistics.* New York: Academic.

Matsuo, Yutaka, Yann LeCun, Maneesh Sahani, Doina Precup, David Silver, Masashi Sugiyama, Eiji Uchibe, and Jun Morimoto. 2022. "Deep Learning, Reinforcement Learning, and World Models." *Neural Networks* 152:267–75. Accessed August 29, 2023. https://doi.org/10.1016/j.neunet.2022.03.037.

Maturana, Humberto, and Francisco Varela. 1980. *Autopoiesis and Cognition: The Realization of the* Living. Dordrecht: Springer.

Merleau-Ponty, Maurice. [1945] 2005. *Phénoménologie de la perception.* Paris: Gallimard.

Moyal-Sharrock, Danièle. 2013. "Wittgenstein's Razor: The Cutting-Edge of Enacticism." *American Philosophical Quarterly* 50:269–79.

–. 2022. *Certainty in Action: Wittgenstein on Language, Mind and Epistemology*. London: Bloomsbury.

Munakata, Yuko. 2016. "Information Processing Approaches to Development." In *Handbook of Child Psychology: Sixth Edition. Volume Two: Cognition, Perception, and Language*, edited by Deanna Kuhn and Robert Siegler, 426–63. Hoboken, NJ: Wiley.

Mundy, Peter. 2018. "A Review of Joint Attention and Social-Cognitive Brain Systems in Typical Development and Autism Spectrum Disorder." *European Journal of Neuroscience* 47:497–514.

Murphy, Kevin P. 2023. *Probabilistic Machine Learning.* Cambridge, MA: MIT.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83:435–50.

–. 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. Oxford: Oxford University Press.

Neisser, Ulric. 1967. *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Noble, Raymond, and Denis Noble. 2023. *Understanding Living Systems.* Cambridge: Cambridge University Press.

OpenAI. "Pioneering research on the path to AGI." Accessed August 9, 2023. https://openai.com/research/overview.

Orhan, Emin, Vaibhav Gupta, and Brenden Lake. 2020. "Self-supervised Learning Through the Eyes of a Child." *Advances in Neural Information Processing Systems* 33:9960–71.

Pavlick, Ellie. 2023. "Symbols and Grounding in Large Language Models." *Philosophical Transactions. Series A: Mathematical, Physical, and Engineering Sciences* 381:20220041. Accessed August 29, 2023. https://doi.org/10.1098/rsta.2022.0041.

Piantadosi, Steven, and Felix Hill. 2022. "Meaning Without Reference in Large Language Models." Preprint. Accessed August 29, 2023. https://arxiv.org/abs/2208.02957.

Polger, Thomas W., and Lawrence A. Shapiro. 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.

Putnam, Hilary. 1975. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
Rescorla, Michael. 2020. "The Computational Theory of Mind." In *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), edited by Edward Zalta. Accessed August 29, 2023. https://plato.stanford.edu/archives/fall2020/entries/computational-mind/.
Ruiz-Mirazo, Kepa, and Alvaro Moreno. 2004. "Basic Autonomy as a Fundamental Step in the Synthesis of Life." *Artificial Life* 10:235–59.
Russell, Stuart, and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. London: Pearson.
Sag, Matthew. (forthcoming) 2023. "Copyright Safety for Generative AI." *Houston Law Review* 61. Accessed August 29, 2023. http://dx.doi.org/10.2139/ssrn.4438593.
Scruton, Roger. 2019. *On Human Nature*. Princeton, NJ; Oxford: Princeton University Press.
Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
–. 1999. "Chinese Room Argument." In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert Wilson and Frank Keil, 115–16. Cambridge, MA: MIT.
Shannon, Claude Elwood, and Warren Weaver. [1949] 1998. *The Mathematical Theory of Communication*. Chicago, IL: University of Illinois Press.
Sharkey, Noel, and Tom Ziemke. 2001. "Mechanistic Versus Phenomenal Embodiment: Can Robot Embodiment Lead to Strong AI?" *Cognitive Systems Research* 2:251–62.
Steinmann, Jan Juhani. 2020. "Die universalsprachliche Singularität und ihre Tücken." *FZPhTh* 67:170–92.
Stewart, John, Olivier Gapenne, and Ezequiel Di Paolo. 2010. *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge, MA: MIT.
Strickland, Eliza, and Glenn Zorpette. 2023. "The AI Apocalypse: A Scorecard. How Worried Are Top AI Experts About the Threat Posed by Large Language Models Like GPT-4?" *IEEE Spectrum*, June 21, 2023. Accessed August 28, 2023. https://spectrum.ieee.org/artificial-general-intelligence.
Tallis, Raymond. 2004. *Why the Mind Is Not a Computer: A Pocket Lexicon of Neuromythology*. 2nd ed. Exeter: Imprint Academic.
–. 2020. *Seeing Ourselves: Reclaiming Humanity from God and Science*. Newcastle: Agenda.
Taylor, Charles. 2016. *The Language Animal: The Full Shape of the Human Linguistic Capacity*. Cambridge, MA: The Belknap Press of Harvard University Press.
Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. London: Penguin.
Thompson, Evan, and Mog Stapledon. 2009. "Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories." *Topoi* 28:23–30.
Thompson, Evan. 2007. *Mind in Life*. Cambridge, MA: Belknap/Harvard University Press.
Tomasello, Michael, and Michael Farrar. 1986. "Joint Attention and Early Language." *Child Development* 57:1454–63.
Tomasello, Michael. 2014. *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.
Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 59:433–460.

Turner, J. Scott. 2017. *Purpose and Desire: What Makes Something 'Alive' and Why Modern Darwinism Has Failed to Explain It*. New York: Harper Collins.

Van Aken, Betty, Benjamin Winter, Alexander Löser, and Felix Gers. 2020. "VisBERT: Hidden-State Visualizations for Transformers." In *Companion Proceedings of the Web Conference 2020 (WWW '20)*, 207–11. Association for Computing Machinery, New York, NY. Accessed August 29, 2023. https://doi.org/10.1145/3366424.3383542.

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind*. Cambridge, MA: MIT.

Varela, Francisco J. 1997. "Patterns of Life: Intertwining Identity and Cognition." *Brain and Cognition* 34:72–87.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is All you Need." In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, 6000–10. Accessed August 29, 2023. https://arxiv.org/pdf/1706.03762.pdf.

Venturinha, Nuno. 2016. "Moral Epistemology, Interpersonal Indeterminacy and Enactivism." In *Action, Decision-Making and Forms of Life*, edited by Jesús Padilla Gálvez, 109–20. Berlin; Boston: De Gruyter.

Vig, Jesse. 2019. "A Multiscale Visualization of Attention in the Transformer Model." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 37–42. Florence: ACL. Accessed August 29, 2023. https://arxiv.org/abs/1906.05714.

Ward, David, David Silverman, and Mario Villalobos. 2017. "Introduction: The Varieties of Enactivism." *Topoi* 36:365–75. Accessed August 29, 2023. https://doi.org/10.1007/s11245-017-9484-6.

Wheeler, Jonathan. 1990. "Information, Physics, Quantum: The Search for Links." In *Complexity, Entropy, and the Physics of Information*, edited by Wojciech H. Zureck, 3–18. Redwood City, CA: Addison-Wesley.

Wiener, Norbert. 1948. *Cybernetics or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT.

Williams, Rowan. 2014. *Edge of Words: God and the Habits of Language*. London: Bloomsbury.

Wittgenstein, Ludwig. 1969–1975. *On Certainty*, translated by Denis Paul and G. E. M. Anscombe, edited by G. E. M. Anscombe and Georg Henrik von Wright. Oxford: Blackwell.

Wittgenstein, Ludwig. 1961. *Tractatus logico-philosophicus,* edited by Luciano Bazzocchi and P. M. S. Hacker. New York, NY: Anthem Press.

–. 1993. "Cause and Effect: Intuitive Awareness." In *Philosophical Occasions: 1912–1951*, edited by James C. Klagge and Alfred Nordmann, 371–426. Indianapolis: Hackett.

–. 1997. *Philosophical Investigations*. 2nd ed. Oxford: Blackwell.

–. 2005. *The Big Typescript: TS 213*. German-English Scholar's Edition. Hoboken, NJ: Wiley-Blackwell.

Zahavi, Dan. 2015. "You, Me, and We: The Sharing of Emotional Experiences." *Journal of Consciousness Studies* 22:84–101.

Ziemke, Tom. 2016. "The Body of Knowledge: on the Role of the Living Body Grounding Embodied Cognition." *Biosystems* 148:4–11.

–. 2020. "Understanding Robots." *Science Robotics* 5:eabe2987. Accessed August 29, 23. https://doi.org/10.1126/scirobotics.abe2987.
Zlatev, Jordan. 2003. "Meaning = Life (+ Culture): An Outline of a Unified Biocultural Theory of Meaning." *Evolution of Communication* 4:253–96.

Oliver Dürr
University of Zurich (Zurich, CH) / University of Fribourg (Fribourg, CH)
orcid.org/0000-0003-0970-4090

Jan Segessenmann
University of Fribourg (Fribourg, CH)
orcid.org/0000-0001-7754-2578

Jan Juhani Steinmann
Catholique University of Paris (Paris, FR) / University of Vienna (Vienna, AUT)