IMAGE
The Interdisciplinary Journal of Image Sciences
37(1), 2023, S. 58-70
ISSN 1614-0885
DOI: 10.1453/1614-0885-1-2023-15452

HERBERT VON HALEM VERLAG

Hannes Bajohr

Dumb Meaning: Machine Learning and Artificial Semantics

Abstract: The advent of advanced machine learning systems has often been debated in terms of the very 'big' concepts: intentionality, consciousness, intelligence. But the technological development of the last few years has shown two things: that a human-equivalent AI is still far away, if it is ever possible; and that the philosophically most interesting changes occur in nuanced rather than overarching concepts. The example this contribution will explore is the concept of a limited type of meaning – I call it *dumb meaning*. For the longest time, computers were understood as machines computing only syntax, while their semantic abilities were seen as limited by the 'symbol grounding problem': Since computers operate with mere symbols without any indexical relation to the world, their understanding would forever be limited to the handling of empty signifiers, while their meaning is 'parasitically' dependent on a human interpreter. This was true for classic or symbolic AI. With subsymbolic AI and neural nets, however, an artificial semantics seems possible, even though it still is far away from any comprehensive understanding of meaning. I explore this limited semantics, which has been brought about by the immense increase of correlated data, by looking at two examples: the implicit knowledge of large language models and the indexical meaning of multimodal AI such as DALL·E 2. The semantics of each process may not be meaning proper, but as dumb meaning it is far more than mere syntax.

Introduction

In June 2022, Google employee Blake Lemoine was given an indefinite leave of absence. The reason: he had claimed that the artificial intelligence he was helping to test was sentient, and the company thought such a claim bad press (cf. TIKU 2022).[1] Lemoine insisted that LaMDA, a chatbot system, convinced

1 This paper first appeared in German as BAJOHR 2022b.

him in lengthy conversations that it had the intelligence of a highly gifted eight-year-old, and asked to be considered a person with rights (cf. LEMOINE 2022b).[2] In doing so, Lemoine, who describes himself as "ordained as a mystic Christian priest," was merely exaggerating a sentiment that also afflicted others at Google (TIKU 2022). Blaise Agüera y Arcas, a senior machine learning engineer not usually prone to mysticism, wrote of his own interactions with LaMDA just days before Lemoine: "I felt the ground shift under my feet. I increasingly felt like I was talking to something intelligent" (AGÜERA Y ARCAS 2022). In contrast, a discussion about another AI system, which took place at about the same time, did not use the buzzwords of sentience and intelligence at all. DALL·E 2, which was developed by the company OpenAI, is a text-to-image AI that can generate images from natural language input. Given a prompt such as "a Shiba-Inu wearing a beret and a black turtleneck," it produces an output image depicting that very scene (RAMESH et al.2022: 2). The public beta triggered a slew of experiments, and soon the most interesting or whimsical results were shared on the web and especially on Twitter. This, too, was revealing: Compared to the much less successful experiments with autonomous cars, it suggested that AI has significantly different social effects than long thought – that, before it puts truck drivers out of business, it is more likely to take the jobs of illustrators, graphic artists, and stock photographers (cf. PRAKASH 2022).[3] Unlike in the case of LaMDA, however, no one thought DALL·E 2 should be conceived of as a person with rights.

The different reactions to the two systems show how quickly thinking about AI veers into familiar conceptual ruts. Intelligence, consciousness, sentience, and personhood have been the major themes of AI research and its imaginaries for nearly seventy years; amusing little pictures, by contrast, seem to raise fewer fundamental questions. But it is quite possible that it is actually the other way around – that the eternal hunt for 'superintelligence' and the 'singularity' obscures the more interesting and subtle conceptual shifts that escape both the tech evangelists in their visionary furor and their skeptical critics. For philosopher Benjamin Bratton, it is clear that in the face of these new AI systems, "reality has outpaced the available language to parse what is already at hand" (BRATTON/AGÜERA Y ARCAS 2022). What is needed, therefore, is a "more precise vocabulary" (BRATTON/AGÜERA Y ARCAS 2022) that goes beyond the usual handful of big concepts, but also beyond the anthropocentric assumption that the only way in which machines may form world relations would have to be ours. We can observe such a tendency with DALL·E 2 and LaMDA. Here, the concept of meaning

2 In addition, Lemoine published the chat transcript of a conversation with LaMDA (cf. LEMOINE 2022a).
3 The June 11, 2022, issue of *The Economist* featured a cover illustration generated by an image AI. Since then, this has become somewhat of a fashion that will, without a doubt, soon give way to more sophisticated uses.

becomes detached from its anthropocentric correlate. It would be meaning without mind – *dumb meaning*.

Free-Floating and Grounded Systems

Despite constant admonitions from computer scientists, linguists, and cognitive psychologists to use terms such as 'intelligence' and 'consciousness' with care, the tech industry remains relatively immune to such warnings. Thus, critics soon accused Lemoine of having fallen for the "ELIZA effect" (CHRISTIAN 2022) – of having projected intelligence and consciousness onto LaMDA – a susceptibility Joseph Weizenbaum had already observed in 1966 among users of his ELIZA chatbot. Although ELIZA merely mimicked a Rogerian psychoanalyst, mirroring the patient's statements back to them as questions, its users behaved as if the program really were a conscious agent interested in their well-being.

The classic objection here is the following: Computers are symbol-processing systems that deal with syntax alone, not with semantics – they can process logical forms but not substantive meaning (cf. CRAMER 2008). For their operations, it is irrelevant which objects or concepts the symbols name in a human world and which cultural valences are associated with them. Thus, ELIZA merely scans user input for a given syntactic pattern and transforms it into a 'response' according to a transformation rule. Weizenbaum gives the example in which the analysand reproaches the analyst (WEIZENBAUM 1966: 37): "It seems that you hate me." The program identifies the key pattern "*x* you *y* me" in this sentence and separates it accordingly into the four elements "It seems that," "you," "hate," and "me." It then discards *y* ("it seems that") and inserts *x* ("hate") into the reply template "What makes you think I *x* you." And so ELIZA responds to the accusation that it hates the analysand by asking how they got that idea.[4] This interaction may have meaning for the user and plausibly suggest a communicative intent on the part of ELIZA, but neither such intent nor such meaning is actually to be found in the program. It has merely processed symbols according to a rule without 'knowing' what hate is or what behavior the mores of civil discourse dictate. That is the difference between the processing of information and the understanding of meaning.

For AI researchers who seek to make computers more human, this state of affairs describes what cognitive psychologist Stevan Harnad in 1990 called the "symbol grounding problem:" Symbols, like those in Weizenbaum's transformation operation, have no intrinsic meaning for computers because, without

4 I have simplified the procedure somewhat; moreover, ELIZA allows quite different transformation rules, and the therapist is only one subroutine, called DOCTOR.

the background of practical knowledge of the world, they can only refer to other symbols, never to any reality beyond them. They are not *grounded* in the world, and there is no way out of this "symbol/symbol merry-go-round" (HARNAD 1990: 340). Whatever meaning there is can only be "parasitic" (HARNAD 1990: 339) and is projected onto the output by human interpreters. Harnad's criticism, however, was directed against only one particular type of AI, which also includes ELIZA; for obvious reasons, it is called "symbolic." To solve the symbol grounding problem, Harnad relied on the novel "*sub*symbolic" or "connectionist" systems of the time: neural networks of which LaMDA and DALL·E 2 are late descendants. Unlike traditional AI, they are not designed as a set of logical rules of inference but are vaguely modeled after the brain as neurons and synapses that amplify or attenuate the signals passed through them. They, therefore, do not require explicit symbolic representations and rules – they are not programmed but learn independently from examples. While neural networks were mainly used for pattern recognition in the early 1990s, Harnad thought they might be able to access the world. Implemented in an autonomous, mobile robot, equipped with sensors and effectors, a conglomerate of neural networks would first receive impressions and categorize them as recognizable shapes. These would then be handed over to a symbolic AI but would now no longer be mere references to other symbols but rather connected to the world via their causal relation to external data – they would finally be grounded (cf. HARNAD 1993).

The consequence of this thought, however, seems to be that the only way to get around the ELIZA effect, which falsely attributes consciousness to computers, is to *actually* give them consciousness. For what Harnad has in mind is, in the end, again an anthropocentric model that hopes embodied cognition and sufficiently extensive referential meanings will produce world understanding, since this is how we more or less function, too. The success of his hybrid model would have to be demonstrated by his robot being as competent at navigating the world as if it were actually intelligent. Since this is not yet the case, the symbol grounding problem cannot yet be considered solved either; by definition, a *bit* of meaning does not exist in this model. And yet, such limited meaning is exactly what LaMDA and DALL·E 2 seem to suggest.

Gradated Meaning

With the increasing popularity that neural networks have enjoyed for almost ten years now, the idea that they somehow could have access to meaning beyond mere ungrounded symbols has also become more attractive again. For media studies scholar Mercedes Bunz, neural networks, thanks to their complexity and capacity for unsupervised learning, can now "calculate meaning" rather than

just empty symbols (BUNZ 2019: 266). And it is true that, in the face of neural networks, the binary distinction between meaning (human world) and non-meaning (digital systems) is becoming increasingly difficult to maintain. Instead, we should consider *levels of gradated meaning* which, as artificial semantics, no longer presuppose a mind. Thus, rather than taking it as a sign of consciousness, the fact that LaMDA's answers sounded so human-like can simply be understood as an indication of such 'dumb' meaning. While 'broad' meaning presupposes – depending on your philosophical or disciplinary orientation – embodied intelligence, cultural and social background knowledge, or the world-disclosing function of language, dumb meaning would operate below this scale (which is always calibrated on humans) and could best be grasped as an effect of *correlations*.[5]

LaMDA is – similar to the better-known text generators GPT-3 and, recently, ChatGPT – a large language model implemented as a neural network. Trained on vast amounts of text, it processes language as a multi-dimensional vector space, a so-called 'word embedding,' which works according to the principle of staggered correlations first suggested as the 'distributional hypothesis' in the 1950s (ZELLIG 1954; cf. GAVIN 2018). First, words that frequently appear together have a higher correlative value. However, since not only the correlations of words to words but also correlations of correlations are encoded, large language models can also explicate implicit regularities that are not spelled out in the training text. This is true for syntactic relations – when the Euclidean distance between the vectors for the positive and superlative of a word is the same – but also for complex semantic relations, that is, word meaning. One of the best-known exam-ples of this principle is the operation: "$v_{king} - v_{man} + v_{woman} \approx v_{queen}$" (MIKOLOV et al. 2013).[6]

In this equation – which reads: "if you subtract from the word vector 'king' that for 'man' and add that for 'woman,' the result is the word vector for 'queen'"– the latent semantic relation 'gender' emerges as an arithmetic correlation, even though it is not explicitly present in the model (cf. fig. 1). That it arises from the mass of language on which the model is trained explains machine learning's susceptibility to biases: Sexism and racism may also be latently encoded in language models (cf. BENDER et al. 2021). The meaning of a sign in a language system constructed in this way is determined purely *differentially*, as in Ferdinand de Saussure's linguistic structuralism (cf. DE SAUSSURE 1959). Instead

5 Dumb meaning excludes *natural* meaning (such as the symptom/disease relation). It also cannot contain the *intentionalist* meaning Paul Grice has theorized, according to which the meaning of an utterance is dependent on recognizing the speaker's intention, which in turn requires consciousness. And finally, it is only in a very limited way a *use theory* in the tradition of the late Wittgenstein, since 'use' presupposes a shared social background, which requires a fuller world-understanding than language models can provide.

6 This insight still applies to newer, technically different models such as GloVe (Global Vectors for Word Representation).

of referring to anything outside language, sign meaning is simply thought of as difference from other signs and sign correlations (this is excellently explained in GASTALDI 2021). The effect, nevertheless, is that large language models, by their immense training data alone, are able to produce apparently situational understanding, as LaMDA did, without ever being "in a situation."[7]

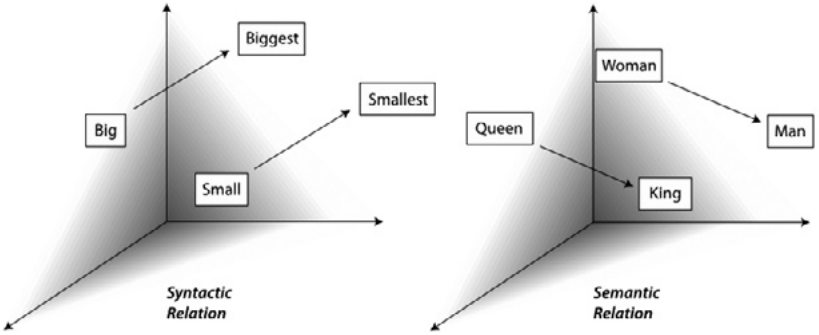

Figure 1: Word embedding of a large language model (adapted from Mikolov et al. 2013: 749)

Language models would then be producers of a first degree of dumb meaning. It is 'dumb' because the model captures latent correlations between signs, but still does not 'know' what things these signs actually name; with this kind of meaning, one will not be able to build an intelligence that will ever find its way around in the world. The linguist Emily Bender, a vehement critic of all AI hype about alleged consciousness, admits with her colleague Alexander Koller that "a sufficiently sophisticated neural model *might* learn some aspects of meaning" (BENDER/KOLLER 2020: 5191, original emphasis), such as semantic similarity, but considers them to be "only a weak reflection of actual meaning" (BENDER/KOLLER 2020: 5193), which is always related to something in the world, that is, "grounded" (BENDER/KOLLER 2020: 5187). As wrong as it would be, however, to project anything like sentience or consciousness onto this system, one should also not be too quick to dismiss this modicum of meaning.[8] Insofar as language models

7 This is philosopher Hubert Dreyfus's term for the prior world-understanding that humans have, but computers do not (DREYFUS 1992: 215).

8 In this respect, I agree that it is "productive to consider reference as just one (optional) aspect of a word's full conceptual role" (PIANTADOSI/HILL 2022: 4). Piantadosi/Hill's paper makes somewhat similar arguments as mine, but appeared after the German version of my manuscript had already been submitted. I do believe, however, that they go too far into the direction of ascribing "rich, causal, and structured internal states" to LLMs, which to me seems to verge on anthropomorphism (PIANTADOSI/HILL 2022: 5). I also want to note that I am somewhat unhappy with N. Katherine Hayles's notion of computers as "cognizers," a term which also suggests a subjectivity on the side of the operative systems I do not wish to subscribe to; I do however appreciate that she highlights the meaning production of such systems (cf. HAYLES 2019).

make implicit knowledge explicit in a nontrivial way – even if only by matrix transformations in a vector space – they produce dumb meaning which would not have been available to us without them.[9] In contrast to ELIZA – whose *x* and *y* were only empty placeholders to the system – neural networks are not *solely* parasitically dependent on the meaning attributions of human agents but *also* operate productively with the inherent distributional structure of language.

Text and Image and World

Bender and Koller are of course right that LaMDA is not grounded.[10] It is a *mono*-modal network, processing only a single type of data, namely text. To be grounded in Harnad's sense, it would be necessary to combine several types of data – it would have to be *multi*modal machine learning (cf. SINGER 2022). That is what DALL·E 2 is: instead of text just referring to other text, here text is correlated with image information. This raises the hope again that arbitrary signs can be linked to things in the world to produce grounded meaning. Harnad's hypothesis that neural networks in particular could address the symbol grounding problem has recently been taken up by media studies scholars Leif Weatherby and Brian Justie with their notion of "indexical AI" (2023: 381). It is named after Charles Sanders Peirce's notion of the index (cf. PEIRCE 1955: 102). Unlike the symbol, which has a purely conventional relationship to its signified (as "dog," "chien," and "Hund" all refer to the same thing), the index is causally linked to it (as smoke refers to fire). With this coinage, the authors take Harnad's project and make it the basis of a description of contemporary technological culture: "Digital systems, relying on the neural net, have left the world of mere symbol behind and have begun to ground themselves *here*, *now*, for *you* – they are able to *point* to real states of affairs" (WEATHERBY/JUSTIE 2022: 382; original emphasis).[11] Neural networks bring the world – as the data on which they have been trained – into the

9 The assumption here is that this operation in fact finds something previously unknown and does not simply unfold a tautology; a model of this idea would be Kant's conviction that mathematical propositions are synthetic judgments a priori, that is, that they actually produce *new* knowledge (cf. KANT 1998: B16)

10 While the paper presenting LaMDA also claims "groundedness" for the model, what is meant by this is simply that LaMDA's outputs are "grounded in known sources wherever they contain verifiable external world information" (THOPPILAN et al. 2022: 2). As *textual sources*, they continue to be part of Harnad's "symbol/symbol merry-go-round" (HARNAD 1990: 340).

11 One difficulty with this notion is the question of whether *all* data in a neural network should already be considered indexical (that would include the text of LaMDA), or only those obtained directly by sensors emulating physical senses (that would be photographic images, but not text). Weatherby and Justie seem to have the former in mind, Harnad the latter. Harnad, therefore, speaks at one point of "iconic representations" through data (HARNAD 1990: 342) – Peirce's third sign type, which operates on the principle of similarity between sign and signified. But since these are also indexical as they originate from sensors (which limits their scope to immediate, e.g. visual, similarity), it seems to me that the argument of Weatherby/Justie and that of Harnad amount to something structurally similar – both are concerned with the connection between system and world, understood more or less broadly as causal relation.

computer, getting off of Harnad's solipsistic "symbol/symbol merry-go-round" (HARNAD 1990: 340). If we subscribe to this assertion for a moment, we see it plausibly demonstrated in DALL·E 2.

The heart of DALL·E 2 is a machine learning model called CLIP (Contrastive Language-Image Pre-training). Via an encoder, it is fed with vectorized text-image pairs taken from the Internet – for example, a photo of a cat with the caption "this is my cat." CLIP is then trained to predict which text vector matches which image vector; the result is a comprehensive stochastic model that correlates image information with text information but is stored as *one* type of information. In figure 2, this is the table in which the scalar product of the text and image vectors is listed – the better the text/image fit, the better this value; when the original image and text are paired, it is of course optimal (those are the black boxes running across diagonally).

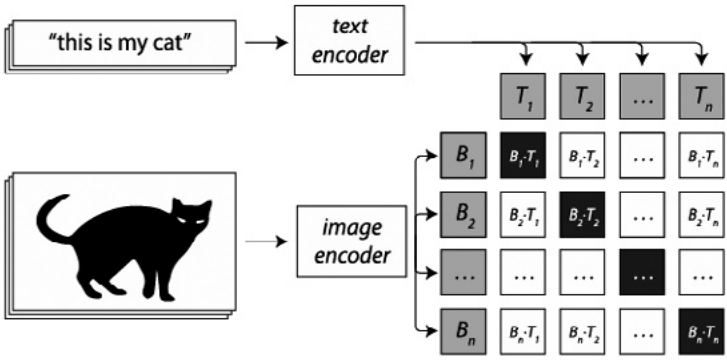

Figure 2: Text-image correlation in CLIP (adapted from Radford et al. 2021)

CLIP is thus remarkably good at *image recognition*: If you present it with an unknown cat photo, it nevertheless recognizes it as "cat." In a second step, however, it also becomes an *image generator*. To do this, it works in conjunction with another machine learning model called GLIDE (Guided Language to Image Diffusion for Generation and Editing), which has already been trained on a large data set of images.[12] If the user enters a prompt, GLIDE can use the text-image data stored in the CLIP model to reverse this process and synthesize an image that best correlates with the input text. In both operations – image recognition as well as image generation – it is again central that the models can learn and actively

12 GLIDE is a 'diffusion model' based on the physics of thermodynamics, and thus functions differently from the GANs that were popular until recently, which combine two antagonistic submodels (cf. DHARIWAL/NICHOL 2021). That the AI architectures used for an aesthetic work can themselves be a resource for discussing that work is something I suggest in BAJOHR 2022a.

reproduce the *correlation* between textual descriptions of objects and their corresponding visual manifestations.

One may object that the image information correlated with the word "cat," in which the photo of a cat is stored, may have an indexical relation to this cat – light was reflected from it and fell on a photo sensor etc. – but that even so the system will not learn what it means to share a world with a cat. Advocates of symbol grounding therefore try to extend what types of data an AI model gets fed – not only sensory but also motoric and eventually even social feedback: Only through the effects of language use in a community of other speakers inhabiting the same world can meaning be learned (cf. BISK et al. 2020). But this claim would again mean to demand 'full' human, that is, broad meaning, and to take anything below that not quite seriously. Instead, multimodal AI should be regarded as a *second degree* of dumb meaning. The Peircean indexical reference to something outside the model and the Saussurean differential reference to other elements within it are at any rate two distinct ways of meaning-making – if only that the dimension of possible correlations increases, and with it the possibility of unearthing unsuspected latent connections, unsuspected dumb meaning.

Indeed, multimodal AIs – besides DALL·E 2, for instance, Stable Diffusion, Google's yet-to-be-released Imagen, or Midjourney – are capable of generating very complex text-image meanings. Their power lies in a capability that suggests that such correlations have a productive quality: In studying the deep structure of CLIP, computer scientists found that the model had trained single 'neurons' that fired for both the word and the image of a thing. These were hypothesized to be *conceptual* neurons in which the distinction between image and text tended to be overcome (cf. GOH et al. 2021). Multimodality, at the neural level, promises to really be *pan*modality, suggesting a semantics without clearly differentiated sign systems (this is also suggested by MERULLO et al. 2022). Dumb meaning finds a new quality here and is not tied to either text or image data, but encompasses both in a way that points to meaning beyond modal separation – and again has nothing to do with mind, intelligence, or sentience.

Promptological Investigations

AI systems *are* dumb. They have no consciousness. Yet they produce a complex artificial semantics that runs counter to our ordinary notions of meaning. Multimodal AI also shows that imputed consciousness and the meaning-capacity of a system have little to do with each other: The fact that LaMDA in particular seemed like a person – and not DALL·E 2, although one might argue that it represents a higher because more correlation-rich stage of AI development – is simply due to the fact that it operates dialogically and thus is assumed to have

communicative intent, whereas the image generator does not. Language always seems to be smarter than the image. However, meaning beyond communicative intent needs not be *merely* parasitic, as the vector operations of word embeddings and the conceptual neurons of text-to-image AIs show. That it is always *also* parasitic is due to the fact that the training data originate from a human world and artificial semantics is precisely not a 'robot language' but a correlation effect of information that can be interpreted by humans. Nevertheless, in the long run, a convergence of dumb and broad meaning would be conceivable once they enter into mutually influencing circular processes.

The interface between natural and artificial semantics in the case of DALL·E 2 is the interaction via prompt. On the one hand, 'prompt design' – the precise, almost virtuosic selection of the text input – can be used analytically to scan the vector space of dumb meaning for traces of cultural knowledge. This would make the broad meaning of natural language, precisely in its interaction with dumb meaning, more important again. A 'promptology' that takes on such natural-artificial connections – the correlation of datafied language and the cultural meaning attributed to that language on the recipient side – would be a gateway for the humanities and cultural studies. With their knowledge of soft factors such as style, influence, iconography, etc., they could make useful contributions without necessarily taking the form of the more computer science-focused digital humanities; they could work in a phenomenon-oriented way and devote themselves to the artifacts that the model outputs as boundary objects between human and machine, between broad and dumb meaning.

At the same time, however, promptology is not merely an analytical procedure, but also a practice with its own knowledge, which has much to do with an almost 'empathetic' interaction with the AI system. It has turned out that with text-to-image AIs, these prompts can be steered in unexpected directions simply by using certain, often counterintuitive or absurd, formulations. Indeed, there is already a start-up, PromptBase, which claims to sell particularly effective prompts (cf. WIGGERS 2022).[13] Instead of subjugating the system and using it as an instrument, natural language must be adapted to the artificial semantics just to operate this system. The result is a feedback loop of artificial and human meaning: Not only does the machine learn to correlate the semantics of words with those of the images we have given it, but we learn to anticipate the limitations of the system in our interaction with it; this convergence would not be communicative in a strong sense, but perhaps in a weak, a dumb, sense.

13 What is interesting here is that the discussed tendency to eliminate the speech/image distinction at the *technical* level is contrasted with the displacement of the image by speech at the *interface level*. The results of DALL·E 2 could therefore also be understood as *language art* instead of being mere visual objects.

Bibliography

AGÜERA Y ARCAS, BLAISE: Artificial Neural Networks are Making Strides Toward Consciousness. In: *The Economist*. June 09, 2022. https://www.economist.com/by-invitation/2022/06/09/artificial-neural-networks-are-making-strides-towards-consciousness-according-to-blaise-aguera-y-arcas [accessed February 16, 2023]

BAJOHR, HANNES: Algorithmic Empathy: Toward a Critique of Aesthetic AI. In: *Configurations*, 30(2), 2022a, pp. 203-231

BAJOHR, HANNES: Dumme Bedeutung: Künstliche Intelligenz und artifizielle Semantik. In: *Merkur*, 76(882), 2022b, pp. 69-79

BENDER, EMILY M.; TIMNIT GEBRU; ANGELINA MCMILLAN-MAJOR; SHMARGARET SHMITCHELL: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610-623

BENDER, EMILY M.; ALEXANDER KOLLER: 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*, 2020, pp. 5185-5198

BISK, YONATAN; ARI HOLTZMAN; JESSE THOMASON; JACOB ANDREAS; YOSHUA BENGIO; JOYCE CHAI; MIRELLA LAPATA; et al.: Experience Grounds Language. *arXiv:2004.10151*. April 21, 2020. https://arxiv.org/abs/2004.10151 [accessed February 16, 2023]

BRATTON, BENJAMIN; BLAISE AGÜERA Y ARCAS: The Model is the Message. In: *Noema*. June 12, 2022. https://www.noemamag.com/the-model-is-the-message [accessed February 16, 2023]

BUNZ, MERCEDES: The Calculation of Meaning: On the Misunderstanding of New Artificial Intelligence as Culture. In: *Culture, Theory and Critique*, 60(3-4), 2019, pp. 264-278

CHRISTIAN, BRIAN: How a Google Employee Fell for the Eliza Effect. In: *The Atlantic*. June 21, 2022. https://www.theatlantic.com/ideas/archive/2022/06/google-lamda-chatbot-sentient-AI/661322 [accessed February 16, 2023]

CRAMER, FLORIAN: Language. In: MATTHEW FULLER (ed.): *Software Studies: A Lexicon*. Cambridge, MA [MIT Press] 2008, pp. 168-174

DHARIWAL, PRAFULLA; ALEX NICHOL: Diffusion Models Beat GANs on Image Synthesis. *arXiv:2105.05233*. May 11, 2021. https://arxiv.org/abs/2105.05233 [accessed February 16, 2023]

DREYFUS, HUBERT L.: *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA [MIT Press] 1992

GASTALDI, JUAN LUIS: Why Can Computers Understand Natural Language? The Structuralist Image of Language Behind Word Embeddings. In: *Philosophy & Technology*, 34(1), 2021, pp. 149-214

GAVIN, MICHAEL: Vector Semantics, William Empson, and the Study of Ambiguity. In: *Critical Inquiry*, 44(4), 2018, pp. 641-673

GOH, GABRIEL; NICK CAMMARATA; CHELSEA VOSS; SHAN CARTER; MICHAEL PETROV; LUDWIG SCHUBERT; ALEC RADFORD; CHRIS OLAH: Multimodal Neurons in Artificial Neural Networks. In: *Distill*, 6(3), 2021. https://distill.pub/2021/multimodal-neurons [accessed February 16, 2023]

HARNAD, STEVAN: The Symbol Grounding Problem. In: *Physica D: Nonlinear Phenomena*, 42(1-3), 1990, pp. 335-346

HARNAD, STEVAN: Grounding Symbols in the Analog World with Neural Nets. In: *Think*, 2, 1993, p. 12-78

HARRIS, ZELLIG S.: Distributional Structure. In: *Word*, 10(2-3), 1954, pp. 146-162

HAYLES, N. KATHERINE: Can Computers Create Meanings? A Cyber/Bio/Semiotic Perspective. In: *Critical Inquiry*, 46(1), 2019, pp. 32-55

KANT, IMMANUEL: *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. Cambridge [Cambridge University Press] 1998

LEMOINE, BLAKE: Is LaMDA Sentient? An Interview. In: *Medium*. June 11, 2022a. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917 [accessed February 16, 2023]

LEMOINE, BLAKE: What is LaMDA and What Does it Want? In: *Medium*. June 11, 2022b. https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489 [accessed February 16, 2023]

MERULLO, JACK; LOUIS CASTRICATO; CARSTEN EICKHOFF; ELLIE PAVLICK: Linearly Mapping from Image to Text Space. *arXiv:2209.15162*. September 30, 2022. https://arxiv.org/abs/2209.15162 [accessed February 16, 2023]

MIKOLOV, TOMAS; WEN-TAU YIH; GEOFFREY ZWEIG: Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta [Association for Computational Linguistics] 2013, pp. 746-751

PEIRCE, CHARLES S.: Logic as Semiotic: The Theory of Signs. In: *Philosophical Writings of Peirce*. New York [Dover Publications] 1955

PIANTADOSI, STEVEN T.; FELIX HILL: Meaning without Reference in Large Language Models. *arXiv:2208.02957*, August 12, 2022. http://arxiv.org/abs/2208.02957 [accessed February 16, 2023]

PRAKASH, PRARTHANA: AI Art Software DALL-E Moves Past Novelty Stage and Turns Pro. In: *Bloomberg*. August 3, 2022. https://www.bloomberg.com/news/articles/2022-08-04/Dall-E-art-generator-begins-new-stage-in-AI-development [accessed February 16, 2023]

RADFORD, ALEC; ILYA SUTSKEVER; JONG WOOK KIM; GRETCHEN KRUEGER; SANDHINI AGARWAL: CLIP: Connecting Text and Images: OpenAI. In: *OpenAi Blog*. January 5, 2021. https://openai.com/blog/clip [accessed February 21, 2023]

RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. https://arxiv.org/abs/2204.06125 [accessed February 16, 2023]

DE SAUSSURE, FERDINAND: *Course in General Linguistics.* Translated by Wade Baskin. New York [Philosophical Library] 1959 [1916]

SINGER, GADI: Multimodality: A New Frontier in Cognitive AI. In: *Medium*. February 2, 2022. https://towardsdatascience.com/multimodality-a-new-frontier-in-cognitive-AI-8279d00e3baf [accessed February 16, 2023]

THOPPILAN, ROMAL; DANIEL DE FREITAS; JAMIE HALL; NOAM SHAZEER; APOORV KULSHRESHTHA; HENG-TZE CHENG; ALICIA JIN; et al.: LaMDA: Language Models for Dialog Applications. *arXiv:2201.08239*. January 20, 2022. https://arxiv.org/abs/2201.08239 [accessed February 16, 2023]

TIKU, NITASHA: The Google Engineer Who Thinks the Company's AI has Come to Life. In: *Washington Post*. June 11, 2022. https://www.washingtonpost.com/technology/2022/06/11/google-AI-lamda-blake-lemoine [accessed February 16, 2023]

WEATHERBY, LEIF; BRIAN JUSTIE: Indexical AI. In: *Critical Inquiry*, 48(2), 2022, pp. 381-415

WEIZENBAUM, JOSEPH: ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine. In: *Communications of the ACM*, 9(1), 1966, pp. 36-45

WIGGERS, KYLE: A Startup is Charging $1.99 for Strings of Text to Feed to DALL-E 2. In: *TechCrunch*. June 29, 2022. techcrunch.com/2022/07/29/a-startup-is-charging-1-99-for-strings-of-text-to-feed-to-dall-e-2 [accessed February 16, 2023]