[Prepublication version, under review. Forthcoming in Robert Lane, Ed., *Pragmatism Revisited* (Cambridge UP). Do not cite without express permission of the author.]

Peirce and Generative AI

Catherine Legg, Deakin University

As a famously far-sighted philosopher, in the late 1800s Charles Peirce was already turning his mind to the topic of artificial intelligence (AI). In an 1887 paper entitled "Logical Machines" he wrote, "Precisely how much of the business of thinking a machine could possibly be made to perform, and what part of it must be left for the living mind, is a question not without conceivable practical importance" (Peirce 1887: 165). He discussed certain mechanical logical machines that had already been developed by William Stanley Jevons in the UK and Allan Marquand in the US – Marquand being one of his students at Johns Hopkins University – and noted that these machines required human intervention to perform each reasoning step. As such, he suggested, logical machine engineers should next try to develop a reasoning equivalent of the Jacquard loom, which executes pre-stored and arbitrarily complex weaving patterns (Peirce 1887: 170). He thereby elegantly anticipated the concept of a computer program, particularly as the Jacquard loom stores its patterns in punch cards, which directly inspired their use to store programs in early computer systems. Moreover, Ken Ketner has plausibly conjectured that Peirce was the author of the first known design for electronic computer circuits composed of 'AND' and 'OR' gates, which was later discovered in Marquand's papers (Ketner et al 1984).

These Peircean insights concerning AI are already reasonably well-known and discussed amongst Peirce scholars.[1] But AI technology has recently crossed a significant threshold with the development of *large language models* (henceforth: LLMs). These machine learning applications absorb massive data sets of human language usage into a deep neural net structure, after which they can fluently generate analogous new texts across numerous knowledge domains and writing genres. This development has become known as *generative AI* (genAI). The societal implications of this newly empowered AI have already proven immense, as Shannon Vallor writes:

> The accelerating spread of commercially viable artificial intelligence is quickly transforming nearly every economic, cultural, and political domain of human activity, from finance and transportation to healthcare and warfare. AI tools are being used to assess loan risk, identify financial fraud, diagnose cancers, evaluate and rank job applicants, write texts, make art, debug code, discover new drug compounds, pilot autonomous vehicles and weapons, and choose a spouse – to name just a few of AI's most well-known and widely discussed applications. (Vallor 2024: 15)

Here I bring Peirce's thought to bear on this exciting and somewhat overwhelming new face of AI. I argue that not only can Peirce's thought help us understand the limitations of genAI applications as 'cognitive helpmeets', it can advise how they might be used most productively. I will draw particularly on Peirce's distinctive *pragmatist epistemology*, which scaffolds what Gili and Maddalena have usefully dubbed a *rich relational realism* (Gili and Maddalena 2022)[2]. Such a realism requires that if our concepts are to be meaningful we must continually test their practical consequences in processes of inquiry that are open-ended to future times and participants. Peirce's realist epistemology also rests on, and is profoundly structured by, a triadic semiotics which does not simply traffic in *symbols* (which are arguably LLMs' strong suit, more on this below). It also explicitly *indexes* the world through

[1] See for instance (Skagestad 1999; 1996; Fetzer 2001; 2004; Steiner 2013).
[2] See also (Maddalena 2017; Lane 2017).

unmediated existential relations between certain signs and their objects, and structures our understanding through *iconic* (nondiscursive, structural) schemata. My discussion will develop Peircean ideas concerning genAI's current capabilities first with regard to *meaning*, then *knowledge and truth*.

1. GenAI and Meaning

To what degree, and in what way, might genAI be said to understand the meanings of human concepts? In the framework of Peirce's semiotics, this becomes a question about the possibility and nature of *artificial sign interpretation*. I will begin this section by drawing a broad contrast between a Cartesian "private and static" idea-based conception of meaning and a Peircean "public and dynamic" semiotic alternative. I will show how many early AI engineers mistakenly drew on Cartesian philosophy in building their applications, which performed poorly, and show how Peirce's semiotics enables us to analyse such efforts as a doomed attempt to 'pre-process interpretation', which can only produce static, 'dead' signs. I will then consider an argument that relevant Peircean lessons have been grasped in the building of LLMs, insofar as they learn the meanings of terms in text corpora without prior, hand-coded definitions. But I will then argue that today's LLMs do not yet perform full 'artificial sign interpretation', because their learned associations between terms, however rich and fine-grained, are insufficiently scaffolded by robust indices to real-world objects and also insufficiently disciplined by iconic structures – most crucially, logical form.

Cartesian Static Meaning

A 'Cartesian model of meaning' is a very broad concept, intended to encapsulate the entire modern era in philosophy, which was of course diverse and riven with internal argument. Yet this broad purview helps us conceptualize the radical nature of Peirce's departure from it.[3] Most fundamentally, the Cartesian model understands the meaning of any given sign as *determined by the intention of its producer*, which has two key features, philosophically speaking. Firstly, the intention is *private*, located somehow "in" the sign-producer's head. The point of this somewhat distracting spatial terminology is that in principle, only a sign's producer has access to its true meaning. Secondly, the intention is *incorrigible*. The signs that I produce mean all and only what I intend them to.

These claims are visible in Descartes' discussion of *ideas* in his *Meditations* (Descartes 1996). In Meditation III, he states that we only have direct epistemic access to our ideas, because the things which exist in the world are ontologically quite separate from the ideas which accurately or falsely represent them. Given that a thinker's mind is so separate from the "external" world (Descartes even claims to coherently doubt whether the latter exists), he must have direct knowledge of what his ideas mean or he can know nothing at all. Error is possible, but not about what our ideas *mean*, only about the way we assemble them, or insofar as we assume that they resemble reality:

> When ideas are considered solely in themselves and not taken to be connected to anything else, they can't be false; for whether it is a goat that I am imagining or a chimera, either way it is true that I do imagine it…..the only kind of thought where I must watch out for mistakes—are judgments. And the mistake they most commonly involve is to judge that my ideas resemble things outside me. (Descartes 1996[1647]: III, 10)

Although Descartes' modern successors embraced a more naturalistic philosophy which abandoned mind-world dualism, they retained Descartes' concept of the private, incorrigible idea as the basic unit of meaning. Thus Locke famously opined, "[W]ords, in their primary or immediate signification, stand

[3] Moreover, in much of his early work Peirce himself defined his developing pragmatism against 'Cartesianism'.

for nothing but the ideas in the mind of him that uses them…" (Locke 1994: 3, II, ii). Let us now consider a very different framework.

Peircean Dynamic Meaning

Although Peirce's semiotics is frequently treated as forbiddingly complex, its central idea that semiosis is constituted by irreducibly triadic relations possesses a certain streamlined elegance. Firstly, a Peircean sign is itself a triadic relation, composed not merely of 'word and object', but also an *interpretant* which consists in further uses of the same sign to represent the same object. As Peirce notes, "a sign is not a sign unless it translates itself into another sign in which it is more fully developed" (Peirce 1931–1958, 5.594, 1903). Imagine that I identify a new insect species, which looks like a Christmas beetle, but bright pink. I decide to name it *Lamprima roseata*. This new name will not become a genuine sign unless others pick it up and use it to refer to similar beetles. Thus this model of signification effectively analyses intelligibility as repetition, thereby theorizing signs as special kinds of *habits*.[4] In the Cartesian framework, to really know what a sign means, one would need to 'read' the ideas of its producer, which is rendered impossible in principle. By contrast, Peirce understands a sign's meaning as *public*, determined by its usage across an entire community which is open-ended to future times and persons.

It is important to note how ongoing interpretation may augment or even alter the meaning of a given sign. A classic example is the word "atom" as used by Democritus, and today.[5] In ancient Greek, "a-tom" meant indivisible, but of course we have now "split the atom". Yet in some sense we are arguably still talking about the same things Democritus was, and the transition from ancient to present meaning was a continuous series of shifts rather than any full semantic rupture. Thus, by contrast to the Cartesian framework, we now have a public, future-directed, indefinitely corrigible account of meaning. As Peirce famously writes:

> [N]o present actual thought (which is a mere feeling) has any meaning, any intellectual value; for this lies not in what is actually thought, but in what this thought may be connected with in representation by subsequent thoughts; so that the meaning of a thought is altogether something virtual (Peirce 1931–1958: 5.289, 1868).

Peirce also distinguished three *kinds* of signs by the way they denote their objects. An *iconic sign* signifies its object by resembling it – a map of Australia represents the country by having the same shape. A key Peircean definition of iconic signification is that the sign's "parts are related in the same way that the objects represented by those parts are themselves related" (Peirce W5: 164-5, 1885). This shows that Peircean iconicity constitutes a "structural resemblance" which is broader than the popular idea of an 'icon' as some kind of picture.[6] Meanwhile, an *indexical sign* signifies its object through some unmediated existential connection, for instance, a pointing finger which directly indicates a place. Although such co-location is a fertile source of indexical signs, it is not the only one. Causation is another – yellow light in the sky can indicate a distant bushfire. Finally, a *symbolic sign* signifies its object through an arbitrary convention or rule. Prime examples are English words such as 'city', but there are also biological examples which rely on 'natural rules' which receive evolutionary responses from other organisms (e.g. jellyfish which flash red when disturbed).

---

[4] This part of the section draws on my previous analyses (Legg 2005; Legg & Black 2020: 2279)

[5] This example, and others, are discussed at (Peirce 1931–1958: 7.587, 1867).

[6] I explore this idea further in (Legg 2008; 2012).

These three sign-kinds have quite distinct functional roles.[7] Symbols, due to the learned repetition of their defining conventions, create cognitive habits which constitute general concepts capable of carrying our knowledge into the future and interconnecting its parts. Indices, due to the brute actuality of their pointing function, connect our knowledge with particular worldly objects, which can challenge and further shape that knowledge. Meanwhile, icons' structural features enable them to vividly depict objects which may or may not exist, which enables us to think modally. Peirce summarises:

> The value of an icon consists in its exhibiting the features of a state of things regarded as if it were purely imaginary. The value of an index is that it assures us of positive fact. The value of a symbol is that it serves to make thought and conduct rational and enables us to predict the future. (Peirce 1931–1958: 4.448, 1903)

Peirce also explains how the three sign-types work together to create intelligible discourse. In the background lies his "experimentalists' view of assertion" (Peirce 1931–1958: 5.411, 1905), which analyses all thought as inquiry, understood as the curation of beliefs which stably meet experience. As such, my initial baptism of a beetle in front of me as *Lamprima roseata* counts as an indexical sign which existentially connects me to a new aspect of reality. If I begin to describe properties of the new species, each of my descriptors (e.g. 'bright pink', 'mating at dusk') will be an iconic sign, likely directly drawn from my perceptual experience. I attach these icons to my new index to create judgments in propositional form (e.g. "This *Lamprima roseata* is bright pink." "This *Lamprima roseata* mates at dusk.")

These judgments generate expectations that further instances of the species will be relevantly similar, and I thereby begin to generalize by expecting similar icons to apply to similar indices (direct encounters with beetle conspecifics) across time. As we noted above that ongoing interpretation may augment or even alter the meaning of a given sign, such further encounters will complexify my initial iconic schemata *a posteriori*. As I encounter genetically variant instances of the species, at different life stages, I develop a more general sense of their characteristic colour and behaviour. In this way, my initial pictorial impressions of a single beetle begin to grow into general symbolic predications of the entire species. This transition whereby repeated attribution of pictures (icons) across an ever-widening range and variety of instances generates general predicates (symbols) constitutes Peirce's distinctive account of concept-generation.[8]

But my localised investigations are only the beginning, if my sign is to 'launch' into its own semiotic destiny. This will occur if the species is studied by others, and my observations integrated with current biological knowledge to generate significant testable predictions (e.g. "As *Lamprima roseata* belongs to the order Coleoptera, it evolved during the Paleozoic Era"). This further study generates a rich fabric of further symbolic generalizations which are connected by association through my new name (e.g. "*Lamprima roseata* shapes vary according to temperature, with the Tasmanian variant the longest and leanest." "*Lamprima roseata* typically mate at dusk, but on cloudy nights they don't mate at all"). But it is important to note how this web of term-based associative relations *is also structured and disciplined by a network of implicit logical relations*. Understanding these relations enables *Lamprima roseata*'s inquirers to avoid many errors and mis-steps. For instance, everyone knows that if the beetle is pink all over then it is not green, and if it mates at dusk then it does not reproduce asexually. A

[7] Here it is interesting to compare mainstream philosophy of language's focus on developing a univocal account of signification, which has arguably created warring tribes of icon-ish, indexical and symbol-like theories, all subject to seemingly endless counter-examples from their rivals.

[8] I outline a much more detailed account of this process in (Legg 2022).

further essential feature in the background of this story is the existence of a community with genuine interest in the new species, and the motivation to learn more about it. I will now trace the contours of the philosophical ideas explored so far in some recent history of AI engineering.

'Cartesian' Early AI

Early AI researchers assumed *cognitivism*, which holds that thinking consists in information-processing over a discrete and abstract set of internal symbols. Cognitivism makes many Cartesian claims.[9] Its description of the relevant symbols as 'internal' and 'discrete' signals Cartesian privacy. Its description of them as 'abstract' signals Cartesian dualism. In computer programming, such dualism nicely maps onto a hardware-software distinction which assumes that the particular computer architecture deployed to 'run' a program does not in any way colour the information it encapsulates, just as Descartes imagined the nature of his body to be irrelevant to the ideas in his mind. As Tom Froese helpfully glosses:

> …the mind is conceptualized as a digital computer and cognition is viewed as fundamentally distinct from the embodied action of an autonomous agent that is situated within the continuous dynamics of its environment. (Froese 2007: 4)

These assumptions found expression in the *Physical Symbol Systems Hypothesis*, which holds that processing structures of symbols is necessary and sufficient for "general intelligent action" (Newell & Simon 1976: 116).

Accordingly, from the late 1950s through to the 1980s AI engineers endeavoured to build systems which would define authoritative, unambiguous, meaning intentions for their symbols, through hand-coded facts and rules (held in 'frames'), reasoned over using deductive logic. These applications are now frequently referred to as "Good Old-Fashioned AI" (GOFAI). Although their outputs were both predictable and explainable, their hand-coded facts and rules required arduous efforts by highly trained and specialized 'knowledge engineers'. The applications were also unable to generalise to new cases, or deal with unstructured environments. Scalability presented a further major challenge, as the inferential tractability of the systems' reasoning was a major issue even with small trial applications (Dreyfus 1992: 91-151; Wheeler 2005; Cantwell Smith 2019: 23-41). This led to increasing efforts to centralise shareable general-purpose knowledge bases, or *formal ontologies*, which would codify the most fundamental concepts or "categories" pertaining to any knowledge domain. The most systematic and well-resourced effort was arguably the long-running **CYC** project (Lenat & Feigenbaum 1991), but it also failed to make real headway (Cantwell Smith 1991; Cantwell Smith 2019: 37), despite always predicting that after just 5 more years, genuine progress would emerge.[10]

Meanwhile other, more grass-roots, applications began to enjoy rapid, seemingly inexorable, uptake. An outstanding example is the **World Wide Web** (WWW), which offered clear and simple

[9] The Cartesian background to cognitivism has been extensively explored by philosophically-trained AI commentators. For instance, Brian Cantwell Smith summarises four "vaguely Cartesian assumptions" of GOFAI: i) the essence of intelligence is thought, ii) the ideal model of thought is logical inference, iii) perception is at a lower level than thought, and will not be conceptually demanding, iv) the ontology of the world is discrete, well-defined, mesoscale objects standing in unambiguous relations (Cantwell Smith 2019: 7-8). See also the early chapters of (Wheeler 2005). Following the landmark work of (Dreyfus 1992), most commentators seeking an alternative to Cartesianism have turned to Heidegger for a 'hermeneutic critique' of GOFAI, and new ideas about the way forward. It's an interesting question whether a Peircean 'semiotic critique' would have been a better choice. I hope to explore this in more detail in future.

[10] In 2023, project leaders suggested that CYC should be grafted onto the newly emerged ChatGPT, to facilitate a shift "from generative AI to trustworthy AI" (Lenat & Marcus 2023).

protocols for assigning each Web resource a unique 'location' (URL), enabling anyone with server space to upload resources instantly available worldwide (for better or worse). Unsurprisingly, many AI engineers attempted to graft GOFAI onto the WWW in the form of the so-called **Semantic Web**, which was marketed as the natural next stage of the WWW, replacing a 'web of links' with a 'web of meaning'. But rather than the WWW reviving GOFAI, GOFAI arguably sank the Semantic Web (Legg 2007; 2013).

I have previously drawn on Peirce's semiotics to analyse GOFAI's approach of hand-coding facts and rules as an attempt to "preprocess" computer applications' interpretations of the meanings of signs. I argued that Peirce's account of the interpretant shows how, precisely in its foreclosing of further meaning development, this approach can only 'kill' signs, and this is why these applications kept 'failing to launch', despite lavish investment in time and energy worldwide:

> What we have seen is a series of attempts to create *ex nihilo* the meaning of signs on the Web *via* a set of antecedent definitions. Arguably this misunderstands what it is for something to have meaning…[T]he lifeblood of meaning-creation is continued mediation of the sign's object to minds *via* specific uses of the sign in specific contexts for specific purposes. Cartesian dualism, with its idealized pregiven meaning postulated in the sign-user's head, misses this. From a Peircean perspective, the mere fact that [these projects] are not widely used *is* the key argument against their having real significance. (Legg 2013: 134-5)

A better approach, I suggested, would be to "build applications that allow interpretants to freely *grow*, within whatever communities choose to use them….[then] *harvest* those interpretants to produce further interpretants that are possessed of genuine added semiotic value" (Legg 2013: 135). I noted that a significant shift in this direction had been taken by Google with its **PageRank** algorithm, which automatically determines the importance of a given webpage relative to the rest of the Web. The algorithm works by counting each hyperlink *to* that page as a vote of support, but its true power lies in how it weights votes by recursively applying its own ranking system, granting important pages a larger vote on which pages are important. Considering hyperlinks as signs, we can understand the PageRank algorithm as deriving an interpretation from them – the importance of a given webpage – which none of the hyperlinks' human creators ever explicitly intended. An important source of the algorithm's success was its access to unprecedented quantities of data, as explained in 2009 by three Google scientists:

> In many cases there appears to be a threshold of sufficient data. For example, James Hays and Alexei A. Efros addressed the task of scene completion…With a corpus of thousands of photos, the results were poor. But once they accumulated millions of photos, the same algorithm performed quite well. (Halevy et al 2009: 9)

Such observations have even inspired speculation that we humans will shortly no longer need to inquire at all. In a 2008 think-piece entitled "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Chris Anderson provocatively describes how new species were allegedly discovered entirely automatically by a scientist named J. Craig Venter practicing "shot-gun gene sequencing":

> In 2003, [Venter] started sequencing much of the ocean, retracing the voyage of Captain Cook. And in 2005 he started sequencing the air. In the process, he discovered thousands of previously unknown species of bacteria and other life-forms…Venter can tell you almost nothing about the species he found. He doesn't know what they look like, how they live…All he has is a statistical blip — a unique sequence that, being unlike any other sequence in the database, must represent a new species. (Anderson 2008)

I will further discuss this 'discovery' below.

Semantics in LLMs

Let us call this process of mining data at scale in order to derive an interpretation that is 'unintended' (in the Cartesian sense) the Automated Interpretant Strategy. It has been developed much further since the early 2000s, and has been crucial to the success of LLMs. I will now explain some intermediate developments. An important step towards LLMs was the representation of terms in text corpora purely mathematically as vectors in a multidimensional space, based on their surrounding terms (or 'embeddings'). A landmark application was **Word2Vec** (Mikolov et al 2023), which uses a high-dimension similarity measure (essentially, the cosine function between vectors) to automatically judge semantic similarity between terms, achieving results previously only attainable by humans. One may gauge the technology's semantic sophistication from its capacity for analogical reasoning. For instance, taking the vector for *king*, subtracting the vector for *man* and adding the vector for *woman* yields a result very close to the vector for *queen*.[11] Here, once again, we see aspects of human sign-use (contiguous word placement in texts) 'mined' to generate useful interpretation (overall semantic similarities) which the human authors never explicitly intended.

A second important step forward was the development of *deep neural networks*, which arrange artificial neural connections in multiple layers to enable much more nuanced machine learning. Combining these two innovations produced *transformer* architecture, and its key tool of *self-attention*. Here, for each term in a given text, the model 'attends to' the terms nearby, encoding them as a kind of context cloud on the term itself. For example, the term *bank* in, "The bank is near the river" embeds the term *river*, thereby building a geographically-oriented context, while the term *bank* in, "The bank approved the loan" embeds the term *loan*, thereby building a financially-oriented one. The immense power of transformer architecture lies in how each network layer re-iterates the self-attention process, thereby embedding contexts containing previously embedded contexts onto each term. Such exponential complexity enables these networks to better capture and reproduce the rich structures that constitute grammatically correct human sentences, and this enabled Word2Vec-style architectures to generate answers to a 'prompt' with impressive fluency, which led to full genAI in the form of **GPT-3** (Brown et al 2020), then, in late 2022, **ChatGPT**.[12] Now, in many contexts, it appears as though we can communicate with LLMs equally fluently as with our fellow humans.

Let us now consider our question to what degree, and in what way, LLMs may justly be described as understanding the meaning of human concepts. An interesting array of responses to this question is emerging from a variety of disciplines. In 2017, psychologist Sudeep Bhatia offered an initial enthusiastic 'behaviorist defence', arguing that Word2Vec replicates human performance in well-known analogical and heuristic reasoning tasks, such as the famous question whether it is more likely that 'single, outspoken, and very bright' Linda is a bank teller or a feminist bank teller (Bhatia 2017)[13]. Some more recent empirical work seems more inconclusive, though. For instance, computer scientists (Yang et al 2023) show that on certain semantic disambiguation tasks (e.g. parsing the term 'old' in the phrase "old teachers' lounge") the performance of ChatGPT is essentially random. A study by computational linguists (Cai et al 2024) concluded that out of 12 'psycholinguistic tests', ChatGPT exhibited human-like responses in 10 and Vicuna in 7, but some of these tests appear tangential to

[11] See also, "brother is to sister as grandson is to [BLANK]" (Titus 2024: 4).

[12] OpenAI have not released a research paper for this application, but have released a blog post: "ChatGPT: Optimizing Language Models for Dialogue" (OpenAI 2022).

[13] This paper is praised as "the most developed and rigorous behavioral defense of an [LLM's] claim to semantic understanding" (Titus 2024: 5).

semantics proper (e.g. guessing associations between feminine pronouns and word endings, and "whether a non-word refers to a round or spiky shape").

Philosophers are increasingly contributing to these debates, but in my view are struggling to find a clear conceptual foothold from which to draw conclusions. For instance, Patrick Butlin distinguishes a *concept*, which he glosses as "referential content" – presumably extensionally defined – from *cognitive significance,* which appears to be the same notion intensionally defined (concepts which "present themselves to the thinker as 'obviously and incontrovertibly' co-referential" (Butlin 2023: 3081)). He then distinguishes both notions from a *conception*: "a structured body of information connected to a concept" (Butlin 2023: 3081). This is not an easy framework to deploy for our purposes, firstly because a pragmatist perspective maintains no sharp distinction between concepts and conceptions, and secondly because it is quite unclear how to operationalize an extensional-intensional distinction for LLMs. (Who is "the thinker" here? How is "obviously and incontrovertibly" to be measured?) Butlin concludes that whilst LLMs such as GPT-3 cannot understand human language, chatbots such as ChatGPT can, because they possess "agency", and therefore "can represent the familiar objects and properties of human life…because they perform tasks that relate to some of these objects and properties" (Butlin 2023: 3093). This move seems very *ad hoc*, given that an LLM answering a prompt might equally be considered the performance of a task which relates (in some way) to "the familiar objects and properties of human life".

Meanwhile, philosopher Lisa Titus critiques Bhatia's behaviorist defense of LLMs' meaning-understanding as satisfying a Statistical – and therefore not a truly Semantic – Hypothesis to explain LLMs' ability to produce "meaning-semblant behavior" (Titus 2024: 5). She proposes the following definition:

> **Functioning Criterion**. A system with semantic understanding functions in ways that are causally explainable by appeal to the semantic relationships among its states and processes with semantic content, and this functioning typically drives the evolution of these states and processes as well as the system's overt behavior. (Titus 2024: 3)

She claims that although LLMs "carry semantic information", they don't have semantic understanding, because "internal representations are [not] appropriately causally connected to the features they purportedly represent" (Titus 2024: 4). Here we may note Cartesian assumptions in her use of the term "internal representations". Also, her definition of semantic understanding as an appropriate causal relationship with "processes with semantic content" seems to imply the traditional static, reified model of meaning, and raises the question of how to define such processes in a non-circular way.[14]

Peircean semiotic analysis can throw further insight on these issues. Contemporary LLMs have taken the Automated Interpretant Strategy to dizzying new heights, insofar as they can now generate ongoing dialogue with humans. Here we may reference a helpful distinction by James Fetzer between "those marks that are meaningful for use by a system and marks that are meaningful for the users of that system" (Fetzer 2001: 130). Whereas the PageRank algorithm provided an automated interpretant for Google's internal purposes (and with respect to just one key concept, a web page's 'importance'), LLMs are externally facing and usable by humans in ways limited only by our creativity. But this very

[14] Titus basically dodges this question in the paper. One alternative criterion of "semantic content" that she does give is "predicative information", glossed as, "sensitivity to the conditions under which one would be a feminist or a bank teller, which goes beyond sensitivity to statistics of text corpora and into the world" (Titus 2024: 8). But it seems equally difficult to define predication without already defining 'the semantic'.

fluency is now exposing new pitfalls, insofar as word vectors capture associative relationships between terms in extraordinarily fine detail, but little else. *These engineers have thereby skilfully captured a form of symbolicity, but no other sign-kind*. Let's examine from a Peircean perspective the two significant lacks here: indexical and iconic signs. In lacking indexical signs, LLMs lack connection with, and thus accountability to, particular worldly objects. This lack is visible in their behaviour – ChatGPT is notorious for 'confabulating' manifestly false answers to questions. OpenAI, who developed ChatGPT, acknowledge this problem, and their analysis of it is illuminating:

> ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there is currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly... (OpenAI 2022).

Strictly speaking, it's not that ChatGPT is telling fibs, rather, its training is simply to generalise statistical patterns of association from corpora of extant statements ("there is…no source of truth"). Hence, Vallor develops an extended metaphor of an 'AI mirror', and its dangers:

> [T]oday's most advanced AI systems are constructed as immense *mirrors* of human intelligence. They do not think for themselves; instead, they generate complex reflections cast by *our* recorded thoughts, judgments, desires, needs, perceptions, expectations, and imaginings. (Vallor 2024: 2)[15]

A related danger which is fascinating to ponder is so-called "model collapse", in which training LLMs on LLM-generated content – which may have to happen soon, given that LLMs have already 'mined' much of the available free human-generated content – leads the whole system to become garbled in a kind of electronic game of Chinese Whispers. ("We find that indiscriminate use of model-generated content in training causes irreversible defects in the resulting models" (Shumailov et al 2024)). Peirce arguably predicted this insofar as he described the role of indexical signs as like "the hard parts of the body…which hold us stiffly up to the realities" (Peirce 1998: 10).

The lack of iconic signification in LLMs is an even more interesting matter, which has so far largely escaped discussion in the literature. We have noted that iconic signs represent structure non-discursively. One might argue that this is not lacking in LLMs since, as we have seen, transformer architecture embeds an enormous amount of structure onto each term. But I submit that this is not *iconic* structure, in Peirce's sense, insofar as these architectures do not reproduce what Wittgenstein insightfully dubbed "the hardness of the logical must" (Wittgenstein 1956: §49).[16] Thus ChatGPT users have reported 'illogical' exchanges such as the following: "[A]fter ChatGPT told us that Romeo commits suicide at the end of Romeo and Juliet, we asked whether Romeo dies during the play, and it said there was no way to know!" (Lenat & Marcus 2023).[17] Once again, the merely symbolic generalisations across statistical patterns of association that the application is trained to produce are insufficient – this time because they do not enable it to 'recoil absolutely' from statements of the form '*p* and not *p*', which is what is required for logical consistency. From this observation we may infer that the discipline exerted by the 'logical must' is as much pragmatic/moral as mathematical/structural, and Peirce's philosophy

[15] See also (Giannakidou & Mari 2024).

[16] I have explored the resonances of this most suggestive phrase within Peirce's conception of iconicity in a number of publications, including (Legg 2008; 2012).

[17] This is of course only one anecdote, but scientific testing of the specifically logical capabilities of LLMs is in its infancy. One rare example is (Parmar et al 2024), who have developed a set of tests for LLMs' logical capabilities across propositional logic, first-order logic and non-monotonic reasoning: **LogicBench**. In their initial study, ChatGPT scored 48% and GPT-4 64%, on propositional logic.

does indeed teach this. I will develop this point further in the next section, as we examine genAI's semiotic functioning at the propositional level.

2. Generative AI, Knowledge and Truth

Prominent claims are being made for genAI's potential to expand humanity's knowledge and grasp of truth. (Smith 2024). On the educational side, it is envisaged that genAI will be used for "scaling personalized support, diversifying learning materials, enabling timely feedback and innovating assessment methods" (Yan et al 2024; Meyer et al 2024). On the research side, it is claimed, genAI will be used to synthesize new information from vast datasets, create new knowledge *ab initio*, and scaffold unprecedented multilingual research collaborations (Waduge et al 2024). However, many others view genAI as an existential threat to human knowledge production. On the educational side, there are grave issues with student cheating and relatively shallow learning (Yan et al 2024). On the research side, a rapidly growing stream of AI-generated research outputs (Glynn 2024; Dehouche 2021) poses an existential threat to traditional processes for academic advancement and identifying genuine expertise.

At worst, it has been suggested that genAI might destroy *our very concept of truth*, by undermining the means by which we operationalize it, thereby entrenching the 'post-truth age' heralded by some in 2016.[18] For instance, Vallor draws on philosopher Harry Frankfurt's lauded theorization of bullshit as "even more dangerous to our social foundations than lying", by shutting down our capacity and motivation to test for truth:

> In 2018, [Steve] Bannon famously confessed to adopting the strategy of "flooding the zone with shit." To keep the media off the scent of a story, you don't bother to craft a careful lie that needs to be protected. You just drown the public conversation with massive quantities of bullshit, so that no one can even find the story – and if they do, they can't tell the difference between it and fiction. Flood the zone often enough, and people will stop even trying. (Vallor 2024: 120)

Peirce's philosophy shows how we can and must reconceive our notions of 'reality', 'truth' and 'knowledge' to meet these new challenges. Accordingly, in this section I argue that a faulty representationalist realism has led mainstream philosophy into a distorted understanding of knowledge, which has rendered us vulnerable to 'knowledge-semblance' in textual form. Once again, I shall suggest that at least some blame can be assigned to Cartesian philosophy, broadly construed.

Representationalist Realism

Over the past century of philosophical debate, a certain mainstream 'semantics-ontology nexus' has arguably been driven by a *representationalism* tailored to the Correspondence Theory of Truth. Broadly speaking, the Correspondence Theory holds that all true statements map onto discrete, existent worldly 'truthmakers' (David 2015). Thus the truth of "The cat is on the mat" consists in the existence of the cat and the mat, and their relative arrangement. This inculcates the idea that reality (and thus, our knowledge of it) must consist in 'sentence-shaped states of affairs' or nothing. Although there is a recent trend of offering expressive or deflationary accounts of discourses for which truthmakers can be challenging to identify, such as ethics, the perceived need for such accounts merely constitutes the flipside of the same representationalist coin.

Representationalism generates many philosophical problems. Firstly, as just mentioned, truthmakers are difficult to identify for certain discourses, leading them to be (effectively) disparaged

[18] For discussions of 'post-truth' specifically in light of Peirce's philosophy, see (Gili & Maddalena 2022; Legg 2018).

via 'error theories', and 'fictionalisms'. Representationalism also leaves unexplained *how* our language manages to denote discrete, existent truthmakers, when it would appear that linguistic and worldly items are quite unlike one another. Conversely, it leads us to understand truth as fully capturable in a set of propositions (what from a pragmatist perspective we might describe as inquiry's 'outputs'). We thereby fetishise articulate texts, so that when genAI appears, delivering articulate textual outputs on demand, we struggle to negotiate their 'truth-semblance'.

Such issues have led many contemporary intellectuals to reject realism altogether. Richard Rorty has been an influential figure here, tracing realism back to Plato and attributing it extraordinarily far-reaching negative consequences (Rorty 1979). In an interesting recent paper, two philosophers apply similar lessons to the arrival of ChatGPT. Mark Coeckelbergh and David Gunkel argue that in order to effectively navigate ChatGPT's explosion of apparent truth, we must deconstruct "a Platonic distinction between appearance and the real that is at the heart of Western metaphysics and that continues to shape responses to new and emerging technologies" (Coeckelbergh & Gunkel 2024: 2222). This distinction leads us to attempt to reduce the normative to the metaphysical in the form of a "transcendental" (by which they appear to mean univocal) truth. Echoing our earlier critique of Cartesian intention as a basis for meaning, they note with approval the obvious 'death of the author' in genAI's outputs, which they describe as "writing without any breathing, living voice to animate and authorize its sayings. These writings are unauthorized" (Coeckelbergh & Gunkel 2024: 2226). They urge that, rather than following many other commentators in denouncing these new technologies, and trying desperately to build new systems of 'authorization', we should embrace the following relativist conclusion:

> the performances and the materiality of text have and create their own meaning and value… There is no absolute moral truth and no ultimate source of meaning that authorizes what comes to be said. There is the performance and the text, or rather, there are performances and there are writings. (Coeckelberg & Gunkel 2024: 2228)

Accordingly, following Levinas, they rule that "standards of morality, truth, and meaning are socially negotiated" (Coeckelberg & Gunkel 2024). Although these authors offer a rare and insightful socio-historical reckoning with LLMs as a writing technology, unfortunately, like so many contemporary philosophers their argument is vitiated by false dichotomy in assuming that *there is no other form of realism than transcendental representation*. They thereby miss the possibility that truth might be deeply implicated in agency and practice, and yet univocal.

Rich Relational Realism

Peirce's rich relational realism differs from the mainstream semantics-ontology nexus in understanding true theories as *existentially intertwined* with the surrounding world, rather than merely *describing* it in propositional form. Relatedly, it establishes an internal (semiotic) relationship between truth and inquiry. Gili and Maddalena explain the first point well, drawing in pragmatism's focus on agency:

> A [key] characteristic of a rich, relational realism is the overcoming of a dualism between theory and practice, giving birth to a new kind of criticism that is not only based on demonstrative reasoning. This is also a legacy of American pragmatism: human beings grasp reality by performing actions. We perform experiments to understand nature, we produce proofs to understand mathematics, and we write, draw or sculpt to understand human nature. (Gili & Maddalena 2022: 30).

To understand the second point, we must examine Peirce's socialised epistemology, which is operationalized in *communities of inquiry*. For instance, he wrote:

> the very origin of the conception of reality shows that this conception essentially involves the notion of COMMUNITY, without defined limits, and capable of an indefinite increase of knowledge (Peirce 1931–1958: 5.311, 1868).

Peirce presents a future-directed, 'limit concept of truth' (Legg 2014). When we inquire, we become part of a public, truth-seeking community which is indefinitely large, although as individual inquirers we have finite epistemic powers. We must trust this community to – potentially and in the future – know more and better than we ever could as individuals. (Peirce calls such trust 'fallibilism'.) This means that we can have no *criterion* of truth, yet we have a concept of truth which *is not given by any statement in propositional form, but by a vast set of finely interwoven practices*.

We began to explore these practices in our earlier example of the new beetle species, where we saw how, after I name my specimen, a community gathers and seeks to find further instances and study them. Here all three Peircean sign-kinds play mutually supporting roles. My name serves as an indexical anchor to a new aspect of reality, around which a series of iconic signs gathers, as the community observes a range of specimens, over time transmuting the icons into general symbols which are tested and integrated into the scientific record. All of this forms an amusing contrast with Anderson's boast of entirely theory-free discovery of new species through "shot-gun gene sequencing". In Peircean terms, such efforts can consist at most in planting some new indices, but if no community is present to do the further iconic-symbolic work, this will be a literally meaningless exercise.

Once again, Peirce expresses his ideas in a strong challenge to Cartesianism – this time, Descartes' treatment of inquiry as a chain which is only as strong as its weakest link. A much better model, Peirce suggests, is a multiply reinforced, giant cable:

> Philosophy ought to imitate the successful sciences in its methods, so far as to…trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibres may be ever so slender, provided they are sufficiently numerous and intimately connected. (Peirce 1931–1958: 5.265, 1868)

It might be objected that LLMs could serve as terrific tools to support such a vision. Could we not deploy them to swell the community of inquiry's ranks with tireless, detail-oriented, automated researchers, thereby building a richer set of relations with reality, and thus a stronger cable? Perhaps, but I think I have said enough to show that LLMs as currently implemented are not capable of establishing the *right kinds of relations* to count as inquirers in their own right.

Conclusion

In his piece "Logical Machines", Peirce claimed that the task of creating a reasoning machine is actually quite straightforward:

> The secret of all reasoning machines is after all very simple. It is that whatever relation among the objects reasoned about is destined to be the hinge of a ratiocination, that same general relation must be capable of being introduced between certain parts of the machine. (Peirce 1887: 166)

However, the question of the exact nature and scope of these relations is a deep one. For too long, representationalist realisms have painted a misleadingly idealized picture of truth as consisting in sets of propositions (that is, texts, whether actual or ideal) which are claimed to 'correspond to' reality.

Thus, "writing the book of the world" is thought a fitting metaphor for a realist metaphysics (Sider 2013). Questions concerning the *processes* which might create and maintain such remarkable artefacts are put out of frame. GenAI's astounding stream of highly articulate, truth-semblant, yet worthless texts issues a timely challenge to all of us to think further. Peirce's semiotic analysis, by contrast, shows how meaningful concepts, and a grasp of truth, can only occur across multiple cognitive systems who are simultaneously richly related with one another, and with a shared environment in which they continually act and receive feedback, within a broader context of a logical space of reasons.[19] As Peirce noted, "Mere knowledge, though it be systematized, may be a dead memory; while by science we all habitually mean a living and growing body of truth" (Peirce 1931–1958: 6.428, 1893). May clarity about these matters inspire AI engineers to build even more impressive solutions.

REFERENCES:

Anderson, C. (2008). "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired Magazine* **16(7)**, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory, Downloaded Jan 7, 2025.

Bhatia, S., (2017). "Associative Judgment and Vector Space Semantics." *Psychological Review* **124(1)**, 1–20.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S. (2020). "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems* **33**, 1877–1901.

Butlin, P. (2023). "Sharing our Concepts with Machines." *Erkenntnis* **88(7)**, 3079–3095.

Cai, Z., Duan, X., Haslett, D., Wang, S. and Pickering, M. (2024). "Do large language models resemble humans in language use?" In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics: CMCL 2024*. Stroudsberg PA: Association for Computational Linguistics, 37–56.

Cantwell Smith, B. (1991). "The Owl and the Electric Encyclopedia." *Artificial Intelligence* **47(1-3)**, 251–288.

Cantwell Smith, B. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment.* Cambridge MA: MIT Press.

Coeckelbergh, M. and Gunkel, D.J. (2024). "ChatGPT: Deconstructing the Debate and Moving it Forward." *AI and Society* **39(5)**, 2221–2231.

David, M. (2015). "The Correspondence Theory of Truth." *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/truth-correspondence/, Downloaded Jan 7, 2025.

Dehouche, N. (2021). "Plagiarism in the age of Massive Generative Pre-trained Transformers (GPT-3)." *Ethics Sci Environ Polit* **21(March)**, 17–23.

Descartes, René 1996 [1647]. *Meditations on First Philosophy: With Selections from the Objections and Replies*, rev. ed., trans. and ed. John Cottingham. Cambridge University Press.

Dreyfus, H.L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge MA: MIT Press.

Fetzer, J.H., (2004). "The Philosophy of AI and its Critique." In L. Floridi (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell, 117–134.

Fetzer, J.H., (2001). "Signs and Minds: An Introduction to the Theory of Semiotic Systems." *Computers and Cognition: Why Minds are not Machines*, Springer Science & Business Media, 43–71.

[19] Thus, as Steiner wisely notes, the main difference between human and machine intelligence is not embodiment but the potential for socialization (Steiner 2013: 275).

Froese, T. (2007). “On the Role of AI in the Ongoing Paradigm Shift within the Cognitive Sciences.” In M. Lungarella, F. Iida, J. Bongard & R. Pfeifer (eds.), *Proc. of the 50th Anniversary Summit of Artificial Intelligence*, Berlin, Germany: Springer-Verlag, 63–75.

Giannakidou, A. and Mari, A. (2024). “The Human and the Mechanical: Logos, Truthfulness, and ChatGPT.” *arXiv preprint arXiv:2402.01267*.

Gili, G. and Maddalena, G. (2022). “After Post-Truth Communication. A Problematic Return to Reality.” *European Journal of Pragmatism & American Philosophy* **14(1)**. https://journals.openedition.org/ejpap/2795, Downloaded March 3, 2024.

Glynn, A. (2024). “Suspected Undeclared Use of Artificial Intelligence in the Academic Literature: An Analysis of the Academ-AI Dataset.” *arXiv preprint arXiv:2411.15218*.

Halevy, A., Norvig, P., and Pereira, F. (2009). “The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems* (March/April 2009), 8–12.

Ketner, K.L., Stewart, A.F., Marquand, A. and Peirce, C.S. (1984). “The Early History of Computer Design: Charles Sanders Peirce and Marquand’s Logical Machines.” *The Princeton University Library Chronicle* **45(3)**, 187–224.

Lane. R. (2017). *Peirce on Realism and Idealism*. Cambridge: Cambridge University Press.

Legg, C. (2022). “Habits in Perception: A Diachronic Defence of Hyperinferentialism.” In J. Dunham and K. Romdenh-Romluc (eds), *Habit and the History of Philosophy*, London: Routledge, 243–260.

Legg, C. (2018). “The Solution to Poor Opinions Is More Opinions’: Peircean Pragmatist Tactics for the Epistemic Long Game.” In *Post-Truth, Fake News,* eds M. Peters, S. Rider, T. Besley, M. Hyvonen. Cham Singapore: Springer, 43–58.

Legg, C. (2014). “Charles Peirce’s Limit Concept of Truth.” *Philosophy Compass* **9(3)**, 204–213.

Legg, C. (2013). “Peirce, Meaning, and the Semantic Web.” *Semiotica*, **2013(193)**, 119–143.

Legg, C. (2012). “The Hardness of the Iconic Must: Can Peirce's Existential Graphs Assist Modal Epistemology?” *Philosophia Mathematica* **20(1)**, 1–24.

Legg, C. (2008). “The Problem of the Essential Icon.” *American Philosophical Quarterly* **45(3)**, 207–232.

Legg, C. (2007). “Ontologies on the Semantic Web.” *Annual Review of Information Science and Technology* **41**, 407–452.

Legg, C. (2005). “The Meaning of Meaning-Fallibilism.” *Axiomathes* **15(2)**: 293–318.

Legg, C. and Black, J. (2022). “What is Intelligence For? A Peircean Pragmatist Response to the Knowing-how, Knowing-that Debate.” *Erkenntnis* **87(5)**, 2265–2284.

Lenat, D.B. and Marcus, G. (2023). “Getting from Generative AI to Trustworthy AI: What LLMs might learn from CYC.” *arXiv preprint arXiv:2308.04445*.

Lenat, D.B. and Feigenbaum, E.A. (1991). “On the Thresholds of Knowledge.” *Artificial Intelligence* **47(1-3)**, 185–250.

Locke, J. (1994[1685]). *An Essay Concerning Human Understanding*. London: Prometheus Books.

Maddalena, G. (2017). “Scientific and Not Scientistic: The Rich Realism of Pragmatism.” *Rivista di Storia Della Filosofia* **72(3)**, 401–414.

Meyer, J., Jansen, T., Schiller, R., Liebenow, L.W., Steinbach, M., Horbach, A. and Fleckenstein, J. (2024). “Using LLMs to bring evidence-based feedback into the classroom.” *Computers and Education: Artificial Intelligence* **6**, 1–10.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). “Efficient Estimation of Word Representations in Vector Space.” *arXiv preprint arXiv:1301.3781*, *3781*.

Newell, A. and Simon, H.A. (1976). “Computer Science as Empirical Inquiry: Symbols and Search.” *Communications of the ACM* **19(3)**, 113–126.

OpenAI (2022). “ChatGPT: Optimizing Language Models for Dialogue.” https://openai.com/index/chatgpt/, Downloaded Jan 7, 2025.

Parmar, M., Patel, N., Varshney, N., Nakamura, M., Luo, M., Mashetty, S., ... & Baral, C. (2024). "LogicBench: Towards Systematic Evaluation of Logical Reasoning Ability of Large Language Models." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1)*, 13679–13707.

Peirce, C.S. (1931–58). *Collected Papers*, C. Hartshorne, P. Weiss, A. Burks (eds). Cambridge MA: Harvard University Press.

Peirce, C.S. (1887). "Logical Machines." *The American Journal of Psychology* **1(1)**, 165–170.

Peirce, C.S. (1998). *Essential Peirce, Vol. 2: Selected Philosophical Writings (1893–1913)*. Eds. N. Houser and C. Kloesel. Indianapolis: Indiana University Press.

Rorty, R. (1979). *Philosophy and the Mirror of Nature.* Princeton NJ: Princeton University Press.

Santoro, A., Lampinen, A., Mathewson, K., Lillicrap, T. and Raposo, D. (2021). "Symbolic Behaviour in Artificial Intelligence." *arXiv preprint arXiv:2102.03406*.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. and Gal, Y., (2024). "AI models collapse when trained on recursively generated data." *Nature* **631(8022)**, 755–759.

Sider, T. (2013). *Writing the Book of the World*. Oxford: Oxford University Press.

Skagestad, P. (1999). "Peirce's Inkstand as an External Embodiment of Mind." *Transactions of the Charles S. Peirce Society* ***35*****(3)**, 551–561.

Skagestad, P. (1996). "The Mind's Machines: The Turing Machine, the Memex, and the Personal Computer." *Semiotica* **111(3/4)**, 217–244.

Smith, B. (2024). "LLMs and Practical Knowledge: What is Intelligence?" In Kristof Nyiri (ed.), *Electrifying the Future, 11th Budapest Visual Learning Conference, Hungarian Academy of Science*, 19–26.

Steiner, P. (2013). "C.S. Peirce and Artificial Intelligence: Historical Heritage and (New) Theoretical Stakes." In *Philosophy and Theory of Artificial Intelligence*. Berlin, Heidelberg: Springer, 265–276.

Titus, L.M. (2024). "Does ChatGPT have Semantic Understanding? A Problem with the Statistics-of-Occurrence Strategy." *Cognitive Systems Research* **83**, 1–13.

Vallor, S. (2024). *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford: Oxford University Press.

Waduge, A.O., Kulasooriya, W.K.V.J.B., Ranasinghe, R.S.S., Ekanayake, I., Rathnayake, U., and Meddage, D.P.P. (2024). "Navigating the Ethical Landscape of ChatGPT Integration in Scientific Research: Review of Challenges and Recommendations." *Journal of Computational and Cognitive Engineering* **3(4)**, 360–372.

Wheeler, M. (2005). *Reconstructing the Cognitive World: The Next Step*. Cambridge MA: MIT Press.

Wittgenstein, L. (1956). *Remarks on the Foundations of Mathematics*, trs G.E.M. Anscombe, ed. G.H. von Wright and R. Rhees. Oxford: Basil Blackwell.

Yan, L., Greiff, S., Teuber, Z. et al. (2024). Promises and Challenges of Generative Artificial Intelligence for Human Learning. *Nat Hum Behav* **8**, 1839–1850.

Yang, S., Chen, F., Yang, Y. and Zhu, Z. (2023). "A Study on Semantic Understanding of Large Language Models from the Perspective of Ambiguity Resolution." In *Proceedings of the 2023 International Joint Conference on Robotics and Artificial Intelligence*, 165–170.