episteme

ARTICLE

Analysis of Beliefs Acquired from a Conversational AI: Instruments-based Beliefs, Testimony-based Beliefs, and Technology-based Beliefs

Ori Freiman

Digital Society Lab, McMaster University, Canada
Email: freimano@mcmaster.ca

Abstract

Speaking with conversational AIs, technologies whose interfaces enable human-like interaction based on natural language, has become a common phenomenon. During these interactions, people form their beliefs due to the say-so of conversational AIs. In this paper, I consider, and then reject, the concepts of testimony-based beliefs and instrument-based beliefs as suitable for analysis of beliefs acquired from these technologies. I argue that the concept of instrument-based beliefs acknowledges the non-human agency of the source of the belief. However, the analysis focuses on perceiving signs and indicators rather than content expressed in natural language. At the same time, the concept of testimony-based beliefs does refer to natural language propositions, but there is an underlying assumption that the agency of the testifier is human. To fill the lacuna of analyzing belief acquisition from conversational AIs, I suggest a third concept: technology-based beliefs. It acknowledges the non-human agency-status of the originator of the belief. Concurrently, the focus of analysis is on the propositional content that forms the belief. Filling the lacuna enables analysis that considers epistemic, ethical, and social issues of conversational AIs without excluding propositional content or compromising accepted assumptions about the agency of technologies.

1. Introduction

At the beginning of the millennium, philosopher Alvin Goldman asked how traditional epistemological questions are "raised and revisited by developments in the telecommunications technology" (Goldman 2000: 127). He discussed beliefs acquired from these technologies and pointed out that "questions about why and whether those beliefs qualify as knowledge will become more central to our thinking" (142). Some technologies, such as the Internet or smartphones, have not just mundanely added more opportunities for us to choose from – but constantly shape how we live (Waelbers and Briggle

2010). Similarly, more opportunities to acquire knowledge have sprouted. Much of our knowledge, and indeed our decision-making processes, depends on membership in epistemic communities and, no less, on our accompanying technologies.

Conversational AIs are technologies in which we interact with natural language. They are the interface for using products with natural language. For example, they can be personal virtual assistants in which we interact with voice, or chatbots which we interact with text. Users provide textual or voice inputs, which are decoded by the conversational AI. The output is presented to the human in natural language – either by text, or voice.

Commonly found examples of technologies that speak with natural voice are Echo, Alexa (both by Amazon), Google Assistant and Google Now, Cortana (Microsoft), AliGenie (Alibaba), Duer (Baidu), Xiaowei (Tencent), Viv (Samsung), and probably the best known – (Apple's) Siri. Of course, speaking with devices has become a common phenomenon not only with virtual assistants, but also in smart homes and new cars.

Moreover, natural language interfaces also include some kinds of chatbots, that use text-based exchanges for dialogues. The ability to program them to hand out particular information in various ways, 24/7, without a human that immediately operates them, renders them common in commerce, healthcare, education, and more.

At the end of November 2022, OpenAI released their chatbot, called ChatGPT, to the public's usage (OpenAI 2022). This product can help with coding, writing songs, summarizing texts, understanding topics, suggesting creative ideas, authoring texts, and countless more tasks. Unlike previous chatbots, ChatGPT remembers "what was said earlier, explaining and elaborating on its answers, apologizing when it gets things wrong" (Harwell *et al.* 2022). It took five days for 1 million users to adopt this technology (Mollman 2022).[1] The fast adoption has brought ChatGPT and its underlying technology to the forefront of mainstream attention.

From a technical standpoint, ChatGPT and conversational AIs in general, are based on language models. Language models use AI that analyze patterns in large amounts of text to calculate and determine the probability of certain words occurring together. The language models use this information to generate new texts or to understand natural language expressed by humans. Language models can be implemented via statistical and deep learning methods. Statistical language models use statistical techniques, such as N-gram tables, to analyze patterns in text, while deep learning language models use neural networks to learn patterns in data. Deep learning[2] models can handle more complex patterns and larger amounts of data, however require more computational resources (Strubell *et al.* 2019).

There are various types of language models, with examples including Google's BERT (Bidirectional Encoder Representations from Transformers; see Devlin *et al.* 2018),

---

[1]In comparison, it took social media platform Facebook (currently 'Meta') 10 months, and the streaming platform Netflix three years to gain 1 million users (Hurst 2022).

[2]When it comes to deep learning, the word "deep" refers to the use of multiple layers of neural networks that enable the model to learn complex patterns and make accurate predictions. As for "learning", there is a variety of training methods for teaching an AI how to recognize patterns in data and make predictions. Common training methods include supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves providing the model with labeled training data and teaching it to recognize patterns and make predictions; unsupervised learning involves providing unlabeled data and allowing the model to learn patterns on its own; and reinforcement learning involves providing the model with feedback on its performance and rewards for correct predictions (see Russell and Norvig 2021: Ch. 19–22).

Google's LaMDA (Language Model for Dialogue Applications; see Thoppilan *et al.* 2022; Freiman and Geslevich Packin 2022), and OpenAI's GPT-3 (Generative Pre-trained Transformer; see Brown *et al.* 2020; Floridi and Chiriatti 2020; for surveys of conversational AI language models, see Adewumi *et al.* 2022; Fu *et al.* 2022). These models are often used in natural language processing applications such as machine translation, question-answering, and chatbots.

The full scope of chatbots and language models' social, ethical, and philosophical implications is, arguably, currently vaguer than known. They are expected to impact human communication, education, scientific research, politics, legal practice, medical practice, entertainment, and many more aspects of life. While doing so, these technological advancements raise numerous philosophical questions: What would happen when conversational AIs would be able to also take decisions? How can we speak, in social-epistemic terms, about the transfer of misinformation that constitutes beliefs? Can social epistemological theory analyze socio-ethical cases of conversational AIs expressing and spreading sexism (Meaker 2019) and anti-Semitism (Boland 2020), uttering false information (Blake 2019)? Moreover, does anyone hold any responsibility for these devices' truth inputs and outputs? How can language models influence a human's understanding of meaning? What are the ethical considerations that must be considered when creating language models? How do language models expect to impact philosophical notions of truth and knowledge? Speaking with conversational AI is a phenomenon that is expected to grow, and to challenge our current theory of knowledge.

Speaking with devices is perhaps one of the most intuitive interfaces that can be designed for communicating with technologies. Though conversational AIs have become widespread, the say-so of technologies, I argue, cannot be analyzed under existing theoretical terms and assumptions. This paper seeks a concept that will enable an analysis of beliefs acquired from conversational AIs. As will be later argued, this concept must meet at least two demands. First, the non-human agency of conversational AI must be acknowledged. This means that the technical processes of the technology can be included in the analysis. Second, the desired concept must allow future scholars to focus on the content that was delivered. This means that the proposition expressed in natural language can be under epistemic and normative scrutiny.

The structure of the paper is as follows: In section 2, three different approaches to analyzing knowledge from instruments are argued as unsuitable for this mission: the coherentist approach, rational-inductive approach, and the approach of knowledge from indicators. These approaches either focus on perceiving signs and indicators and not natural language content; or emphasize the reliability of the perceptual beliefs and the instrument's reliability rather than the reliability of the content we desire to analyze.

Section 3 considers testimonial theories of knowledge. First, the historical underpinning of the concept of testimony as anthropocentric is given. I show the perspective of two fields: epistemology and sociology. Leading scholars from both fields established the view that a technological artifact cannot constitute a testifier since technologies, unlike people, lack a moral character. Then, this anthropocentric view is shown to be assumed by contemporary testimonial theories of knowledge. The section ends by considering, and rejecting, an alternative and non-anthropocentric view of the concept of testimony.

In section 4, I suggest the concept of technology-based beliefs. This concept is suggested as a complementary concept to instrument-based beliefs and testimony-based

beliefs. This new concept acknowledges the non-human agency of conversational AI, thus avoiding what the anthropocentric notion of testimony cannot. Additionally, the new concept focuses on the delivered natural language content rather than the reliability conditions surrounding the perception. The concept of technology-based beliefs is argued to be a plausible solution to fill the lacuna of acquiring knowledge from conversational AIs.

2. Instrument-Based Beliefs

Traditional methods for acquiring knowledge from instruments focus on epistemological justifications necessary to establish the concept of knowledge. Justifications for knowledge acquired from instruments are insufficient for analyzing knowledge gleaned from technologies that interact with natural language. These methods depend on the reliability of the perceptual beliefs or the instrument's reliability rather than the reliability of the content of the proposition. In addition, they fail to capture the epistemic dependency of an individual on other members of her community.

The field of epistemology traditionally discusses sources of knowledge as either deriving from one's own mind – such as introspection, memory, or reason – or from one's environment through perception and (arguably) the testimony of other people. Therefore, beliefs acquired from instruments and digital technologies are not considered to derive from unique sources of knowledge. However, alternatives do exist. For example, Neges (2018) argues for the view that beliefs acquired from instruments and digital technologies derive from unique sources of knowledge. He suggests that "instrumentation" is a unique epistemic source of belief from instruments, and that this source is not reducible to perception or inference.

Another exception is made by Alvarado (2022a), who argues that AI is a specific kind of instrument that manipulates its content through epistemic operations and is aimed for tasks which are epistemic in nature. Moreover, Alvarado (2022b) makes the case that within the realm of science, computer simulations can be understood as scientific instruments. These exceptions build upon alternatives to traditional approaches to knowledge. An example for such an alternative is the approach of Baird (2004) to knowledge. Baird turns away from the foundational assumption of epistemologists about the concept of knowledge as some sort of a justified true belief. He argues that (some) instruments constitute objective material knowledge, to which he refers as 'thing knowledge' – that embeds and expresses the knowledge of its designers. According to this view, knowledge is not belief-based, but thing-based. Thing-based belief is a radical departure from traditional fundamental assumptions in epistemology (Pitt 2007; Neges 2014; Freiman 2021).

In this paper, I focus on common epistemic traditions, that do not consider beliefs acquired from instruments and digital technologies as deriving from a unique source of knowledge. Instead, as shown in the following sections, they are discussed in terms of perception and inference from sources. The field of epistemology is limited in its ability to offer conceptual tools for the analysis of knowledge whose source is technological.

This is not the case in other related fields. Knowledge production is among the topics of inquiry in some fields, such as STS (e.g., Latour 1986; Collins and Pinch 1993; Lynch 1994; Knorr-Cetina 1999) and post-phenomenology (e.g., Ihde 1991; Verbeek 2005; Olesen 2012). A standard view is that technological artifacts are not neutral intermediaries but actively determine how we construct knowledge (e.g., de Boer et al. 2018). In the subfield of philosophy of science, for example, the epistemic roles

of technological artifacts are acknowledged, but mostly in the context that they enable us to discuss the nature of reality and construct theories (e.g., van Fraassen 1980; Laudan 1981; Hacking 1985; Humphreys 2004; Giere 2006). They do not engage with the mission of how an individual acquires knowledge.

Hereinafter, I deal with accounts of knowledge from instruments. I explore three traditional accounts of knowledge that derive from instruments I refer to as coherentist, rational-inductive, and knowledge from indicators.

2.1. Coherentist Approach to Knowledge from Instruments

One approach to the topic of knowledge and technologies within the field of epistemology is found in the concepts of *instrumental knowledge* and *knowledge from indicators*. According to Lehrer's (1995) *Evaluation Model of Instrumental Knowledge*, to know that *P* through the use of an instrument, a subject must accept its trustworthiness and its output that *P*, as true. This acceptance depends upon having a trustworthy basis for evaluating the belief that *P*, and on being able to defend *P*'s acceptance against possible objections.

Lehrer distinguishes between instruments which aid our senses (such as reading glasses) and instruments that provide information otherwise not available (such as a microscope). In the first case, a person accepts her own senses to be trustworthy; while in the second case, by contrast, a person must accept that the instrument *and* the relevant background theory involved are trustworthy.

Lehrer is considered a coherentist, meaning he denies the notion of basic foundational beliefs. The coherence theory of justification states that a belief is justifiably held if the belief coheres with a set of beliefs. Lehrer's early work on the coherence theory of justification can be distinguished from his later developments: first, Lehrer's (1990) "acceptance system", in which a person needs to accept a belief, and Lehrer's (2000, 2003) "evaluation system" which involves more complex cognitive processes (Olsson 2017: §4). The Evaluation Model of Instrumental Knowledge from 1995 employs both acceptance and evaluation and can be considered an early sign for Lehrer's later (2000, 2003) work.

Acceptance of information from instruments results from an effort to obtain truth and avoid an error. However, it is insufficient (i.e., it can be Gettiered). Only by evaluating beliefs based on the acquired information and knowing how to answer some possible objections about the instrument and its theory can beliefs acquired from instruments be considered as knowledge.

I recognize two problems with Lehrer's account. First, Lehrer spells out the justification condition in terms of the properties of the individual believing subject, rather than considering the social and technical environment. Second, Lehrer does not address the question of the extent to which a knower should know the inner workings of the instrument for defending the acceptance of P against any objections. Defending the acceptance of *P* is the basis for the evaluation of the belief that the instrument itself is trustworthy.

2.2. Rational-Inductive Approach to Knowledge from Instruments

Ernst Sosa's account of knowledge from instruments (Sosa 2006) can give an answer to the evaluation of a belief that an instrument is trustworthy and can solve the second problem posed by Lehrer's account. Phrased differently, Sosa's account addresses the

question of the extent to which a knower should hold knowledge of the inner workings of the instrument. Sosa argues that the notion of testimonial knowledge presupposes the notion of instrumental knowledge. In his account, testimonial knowledge is considered as knowledge that is verbally transmitted from one subject to another using the instrument of language. Because we do not have direct access to another subject's perception, instrumental knowledge, including testimonial knowledge, cannot be reduced to non-instrumental knowledge: "Our access to the minds of others is after all *mediated* by various instruments, and we must trust such media at least implicitly in accessing the testimony all around us" (Sosa 2006: 118, emphasis in original).

A justified belief that an instrument is reliable ultimately derives from relying on our perceptual input. Unlike instruments, our senses differ insofar as we do not need, or cannot have, a rational basis to justify our beliefs. Our senses are "a gift of natural evolution, which provides us with perceptual modules that encapsulate sensory content and reliability in a single package" (Sosa 2006: 122). While we have some kind of a default justification for trusting our senses, we need some rational basis for accepting the output of instruments.

Sosa argues that the basis for accepting an instrument is reliable exists when a subject has an indication that the instrument indicates the outright truth and accepts this indication. The justification for relying on an instrument has an inductive basis: the more a subject uses it, the more she gains support for its reliability (Sosa 2006: 120). Once this rational basis is established, the behavior of relying upon what the instrument delivered is adopted. That is, when an instrument repeatedly produces a truth output, the subject is inclined to incorporate its reliability as an assumption.

Here, too, problems arise: first, like Lehrer, Sosa spells out the conditions for acquiring knowledge from instruments in terms of properties of the individual believing subject, without including social or technical characteristics. Second, Sosa does not explain how a subject acquires indications that the instrument is reliable.

2.3. Knowledge from Indicators

The third example of traditional accounts of acquiring knowledge from instruments comes from Millar's (2009) account of knowledge from indicators. He provides a general approach for explaining how a subject acquires indications, such as those Sosa refers to, and as such can solve the second problem posed by Sosa's account (explaining how a subject acquires indications that the instrument is reliable). Millar's work is grounded in the notion of a subject's successfully exercising her recognitional abilities. While Lehrer assumes that the justification of a subject's belief that an instrument is trustworthy depends on the subject's own evidence for the reliability of the instrument, Millar does not limit this justification only to evidence, but also expands to include the subject's recognitional abilities.[3]

Similar to Sosa's account of knowledge from instruments, Millar's notion of knowledge from indicators fundamentally rests upon perception. For Millar, perceptual knowledge is the exercise of the subject's ability to recognize something she perceives (Millar 2009: 120). When perceiving an indicator, the indication is understood as

---

[3]While Lehrer focuses on possible objections one has in her mind regarding the reliability of the instrument, and is thus an internalist in this sense, Millar focuses on the reliability of the cognitive abilities to acquire indications and therefore can be considered an externalist. For hybrid accounts of internalist and externalist justifications, cf. Henderson *et al.* (2007), Comesaña (2010), and Goldman (2011).

factive, grounded in a causal relation (149). An example is the fuel gauge: I look at my fuel gauge and notice it half-full. It is the quantity of fuel in my car's fuel tank that causes the gauge to indicate that it is indeed half-full. The indication is perceived by my perception.

Millar recognizes and acknowledges the problematic question raised by his account: "Knowledge from indicators is problematic because it seems puzzling that we can be entitled to take indicating phenomena to indicate what they do" (Millar 2009: 162). The causal chain can be considered a black box. If the content of the black box is unknown, the causal chain cannot constitute a part of that which justifies a subject's knowledge as derived from indicators.[4]

2.4. Instrument-based Beliefs and Natural Language Technologies

Traditionally, knowledge from instruments deals with mechanical instruments rather than digital technologies. Indications, such as a green light indicating a device is turned on, a fuel gauge, or a display that says "50 degrees" are different from natural language technologies that speak with words – such as a recording playing the next stop on the subway or speaking with a digital virtual assistant. Physical causality and mechanical explanations describe the indications. At the same time, software code, big data collected from the behavior of many, and algorithms, form the interactions of digital virtual assistants, chatbots, and other technologies that speak. Traditional analysis of knowledge from instruments is not suitable for analyzing natural language and algorithms.

Instrument-based belief is a belief "formed through reliance on an instrument's output or 'read-out'" (Goldberg 2012: 184). Epistemic accounts of analyzing knowledge from instruments mostly assess the reliability of the senses and the reliability of the instrument used for measurement rather than assess the content of the knowledge acquired and its source. For example, an analysis of the green led light on my phone charger might assess the causal chain, perceptual or inferential beliefs, or the reliability of the led light or my eyes to perceive it, but not the propositional content – "the battery is full" or "currently charging". Additionally, these accounts do not capture the epistemic dependency of an individual on the larger technical environment and other people.

Epistemic accounts of analyzing knowledge from instruments are individualistic and suitable for mechanical instruments but not for technologies that speak in natural language. As social creatures, we have always relied on one another to gain knowledge. Since we rely on technologies to acquire knowledge, how else could it be possible to analyze the acquisition of knowledge from technologies that speak in natural language?

3. Testimony-Based Beliefs

Another candidate to help us analyze cases where a person acquires beliefs from a technology that speaks is the concept of testimony within social epistemology. A testimony-based belief is "formed through reliance on another speaker's testimony" (Goldberg 2012: 184). In this section, I first explore the historiography of the concept of 'testimony' – to establish why this concept is not suitable for analyzing knowledge acquired from technologies. I then spell out three assumptions that underlie the current view of

---

[4]See Dahl (2018) for reasons which entitle a subject to acquire knowledge from technologies without inspecting its inner workings.

testimony as incapable of dealing with technologies – having intentions, the capability of being normatively assessed, and taking part in trust relations. I argue that an existing alternative that wishes to treat technologies and humans the same – undermines fundamental assumptions in the field about human agency and is incompatible with the current view of testimony. Lastly, I argue against two possible objections suggesting that the outputs of conversational AIs can be analyzed as a form of group testimony, concluding that testimony-based approach is not suitable to epistemic analysis of beliefs acquired from conversational AI.

3.1. The Social-Philosophical Roots of 'Testimony'

In the coming section, I present the renewed interest of scholars in the concept of testimony during the early 1990s. The origin of the concept of testimony helps us understand why the concept cannot be used today – for analyzing cases of acquiring knowledge from technologies that speak.

Much of what we know, as individuals and as groups, depends on the words of others: "We live in a sea of assertions and little if any of our knowledge would exist without it" (Lipton 1998: 1). Philosophers from all eras have put effort to making sense of how language represents the world, and how we share these representations. Yet, despite the fundamental role of other people's words in knowledge, it is only relatively recently that the concept of testimony has become an object of research. Current research originates from several seminal works.[5] Two of them are Coady's (1992) *Testimony: A Philosophical Study*, and about two years later, Shapin's (1994) *A Social History of Truth: Civility and Science in Seventeenth-Century England*.

Coady's influential monograph discusses philosophical arguments about knowledge acquired from others. Gelfert, in his book dedicated to the concept of testimony, summarizes why Coady's monograph is considered *the* cornerstone of a research program about the say-so of others: "As with most truly influential books in philosophy, perhaps the greatest significance of Coady's book lies in the responses and criticisms it provoked, as well as in the philosophical theories that it inspired others to develop" (Gelfert 2018: 2).

Shapin (1994) studies the production of knowledge in 17th-century England. His cultural-historical research is based upon an argument that the scientific culture of that time was built upon the word of a gentleman. A gentleman's social status was a crucial factor in considerations regarding the question of whom to trust. Unlike common laborers or merchants, gentlemen were not affected by economic pressure, a force that could compromise the ability to tell the truth. Therefore, the question of which testimony we should accept can be answered based on social factors, such as status.

Shapin's work heavily influenced historians of science, sociologists, and specifically sociologists of knowledge – where the concept of testimony became a fundamental theoretical notion. For example, the Strong Programme in the Sociology of Knowledge is committed to the accepted formulation and defense of a theory of knowledge that holds testimony to the principal method by which epistemic communities are formed, and knowledge is generated. This is due to the ability of testimony to establish a social agreement that transforms mere opinion or belief into knowledge (Kusch 2002).

---

[5]Other seminal works include, for example, Fricker and Cooper (1987) who identify testimony as a distinct source of belief; an edited book by Matilal and Chakrabarti (1994); and Fricker's (1995) review of Coady's (1992) book that further sparked social epistemic research on the concept of testimony. For further details, see Gelfert (2018).

Within epistemology, the concept of testimony is used to describe cases in which a testifier asserts a proposition that the receiver of the testimony consequently believes. It is described in terms of a testifier *T*, who testifies a proposition *P*, to a hearer *H* (or a reader, or a receiver of the testimony). Hearer *H*'s belief that *P* can be considered as a testimonial-based belief (Pritchard 2004: 326). Described this way, testimony is the most elementary, yet all-encompassing, concept to describe the knowledge and the justification for knowledge traveling from one agent to another. Testimonies differ in their contents and contexts, and as such, the subfield of the epistemology of testimony grapples with various core issues and debates.[6]

The concept of testimony can be considered a natural candidate for analyzing the case of a person who acquires beliefs by receiving verbal propositions from a device. However, whether or not a device can deliver testimony is debatable: the received view of testimony holds, generally, that only persons can participate in the act of testimony. This view is advocated by most philosophical accounts of testimony (Coady 1992: 268; Lackey 2008: 189).

Similar to the received view of testimony within epistemology, the view that only persons can give testimony is also advocated by leading sociologists (e.g., Collins and Kusch 1998; Bloor 1999; Collins 2010; for further sociological contexts of the concept of testimony, see Neges 2018; Freiman and Miller 2020). As Shapin argues, "in securing our knowledge we rely upon others, and we cannot dispense with that reliance. That means that the relations in which we have and hold our knowledge have a moral character" (Shapin 1994: xxv). Instruments, unlike gentlemen, lack such a character and cannot give testimony.

While both Coady and Shapin focus on the ubiquity of testimony, Coady focuses on the questions which shape a person's justification for accepting the testimony of others. Shapin argues that decisions concerning 'who to believe' are matters of moral and social characteristics (Lipton 1998). The received view in both the fields of epistemology and the sociology of knowledge has developed to reject the possibility of a technology testifier.

3.2. Anthropocentric Assumptions in Testimonial Theories of Knowledge

Having established the historical grounds for the position that a technological artifact cannot give testimony, it is possible to turn to current reasonings that categorically reject this option.

Elsewhere, I (Freiman 2021) recognize the 'anthropocentric view of testimony' as a commonly held view among social epistemologists. The view presupposes that only persons can participate in the act of testimony because only humans, in principle, can be qualified as a testifier. Underlying this view are commonly held assumptions in mainstream social epistemology that a testifier (a) must have intentions to deliver the testimony; (b) be subject to normative assessment; and (c) constitute a putative object in trust relations.

Technologies, arguably, do not have intentions. If testimony requires some kind of intention to deliver a proposition to the recipient of the testimony, then the concept of testimony cannot be used for analyzing knowledge from technologies. For example, Fricker (2015: 179) categorizes "[automated] announcements at railway stations of train

---

[6]Examples for these issues and debates are the reductionism/anti-reductionism debate, transmission/ generation debate, issues of expertise, and the value of knowledge, to name but a few. See, e.g., Green (2008), Carter and Pritchard (2010), Adler (2014 [2006]), and Gelfert (2014, 2018).

times, or automated messages one receives on telephone connections, that sound like a live human voice making statements, but are no such thing" as fake testimony.[7]

Additionally, testifiers must be normatively responsible for what they say. As Fricker (2002: 379) argues, "a teller is normatively responsible for the truth of what she asserts". Since technological artifacts cannot be assigned that responsibility yet, they fail to be considered testifiers. For example, Goldberg (2012: 191) argues that only epistemic agents are "susceptible to full-blooded normative assessment".

Lastly, testifiers must be trusted. Testimonial accounts of knowledge demand that the act of testimony will entail trust between the hearer and the speaker (e.g., Gelfert 2014: §8, 2018: §5). However, according to a commonly accepted approach in the epistemology of trust, only humans can be objects of trust relations, rendering technologies as lacking, in principle, the property of trustworthiness (Miller and Freiman 2020; Freiman 2021). This, yet again anthropocentric view, usually shifts discussions of trust from artifacts to the humans behind the technologies (Pitt 2010: 445). Coeckelbergh (2012) nailed its essence: "direct trust in artefacts is indirect trust in the humans related to the technology".

Moving from technology to AIs, the issue of whether or not, and in what conditions, it is possible to trust AI is extensively discussed (see, e.g., Alvarado 2022a). However, according to the traditional social-epistemic approach that trust entails a human quality that technologies lack, it is controversial to hold the view that AI can, *in principle*, be trustworthy or be an object of trust (Bryson 2018; Rieder *et al.* 2020; Ryan 2020; Freiman 2022). Since according to testimonial theories trust relations cannot be formed with technologies, AIs included, technologies, conversational AIs included, cannot be qualified as testifiers. The category of testimony-based beliefs is not suitable for analyzing beliefs acquired from conversational AIs.

3.3. Against the Argument that Testimony-Based Beliefs are Enough

There are existing approaches to technological artifacts as giving testimonies that contrast the received view. For example, Green (2006, 2008) develops a view according to which technologies, and zombies, can give testimony. Green (2008) argues that (some) of the beliefs that originated from technologies are testimony-based beliefs and that there is no need for a different category. Since beliefs from humans and beliefs from technologies have the same epistemic status and content, are a result of the same cognitive ability by the human hearer, and are experienced the same, the concept of testimony is sufficient.[8]

Green's approach to testimony might solve the problem of acquiring knowledge from technologies that speak. However, this solution comes with a price: The symmetry between technologies and humans runs the risk of undermining the distinction between human and non-human agencies. Specifically, it would associate a non-human with intentions, the ability to be normatively assessed, and as a valid object in trust relations. These are incompatible with the accepted view in social epistemology.

---

[7]While technologies that express automated announcements work completely differently from natural language technologies, they still express the output as propositions.

[8]A similar approach is taken by Smart (2017). Compare both Green (2008) and Smart (2017) with Nickel (2013), who shifts from phenomenological similarities between technologies and humans, to pragmatic considerations of performing functions with speech outputs. However, Nickel's approach is not discussed in terms of the epistemology of testimony.

In addition, it is possible that the outputs of conversational AIs can be analyzed as a form of a group testimony. There are at least two different arguments to make:[9] One sees the conversational AI's outputs as the testimony of the humans whose expressions were used in the training data sets – as a collective, and the second considers the conversational AI's outputs as the testimony of an expert community.

The first possible objection that recognizes the outputs of conversational AIs as a form of a group testimony argues that the corpus of their training data sets can be reduced to humans who expressed the text. While the idea is worth exploring, it is problematic: The epistemology of group testimony would either associate the group testimony to a collective knower or accord the group testimony to the many knowers who make up the group (Lackey 2014; Miller 2015).

While I accept that there are viable possibilities for social structures that can deliver testimony as groups (e.g., commissions, research groups, departments, states, and so forth, see Faulkner 2018), I reject both options regarding conversational AIs: recognizing a technological artifact such as a conversational AI as a collective knower anthropomorphizes it (similar to saying 'Google knows'); and the individual humans, whose texts contributed to the data sets, lack the intention necessary for that act to be considered as giving testimony. Additionally, such a notion places a smokescreen on the ability to normatively analyze the algorithms involved in generating the propositions and those who engineered them – humans and institutions, as responsible and accountable for the product (Freiman and Geslevich Packin 2022). In this case, the notion of group testimony is not suitable for acquiring beliefs from the outputs of conversational AIs.

A second approach to identifying the outputs of AIs as a form of group testimony might derive from a discussion about computer simulations. In Symons and Alvarado's (2019) discussion of what it means to trust the results of a computer simulation, they raise the question of whether computer simulations are sources of expert testimony. While some scholars, as their argument goes, argue that trusting the results of computer simulations poses similarities to trusting the say-so of expert testimony or to trusting perception, they suggest that trust is given to the testimony of expert communities, rather than directly the output of the simulations. To use their example, laypeople trust the judgment of meteorologists, with respect to the models they use to predict if a hurricane is likely to hit their city. While in their example, laypeople do not engage directly with the output of the weather models, in the case of conversational AIs, users engage directly with the output of the technology. The testimony of the expert community is not a notion suitable for analysis of acquiring beliefs from the outputs of conversational AIs.

The lacuna can now be clearly identified: traditional approaches in epistemology and their notion of instrument-based beliefs and approaches in social epistemology and their notion of testimony-based beliefs fail to provide a proper ability to analyze knowledge acquisition from conversational AIs. In the next section, I build upon an existing distinction between instrument-based beliefs and testimony-based beliefs. I suggest adding a new, third, category: technology-based beliefs. Technology-based beliefs acknowledge the non-human agency of the source of knowledge acquired and, at the same time, acknowledge the verbal content of the proposition delivered. Technology-based beliefs address the lacuna of analyzing knowledge acquisition from technologies that speak in natural language.

---

[9]I thank an anonymous reviewer for raising both of these objections.

4. Technology-Based Beliefs

4.1. Anthropocentricity: Revolutionizing Social Epistemology or Maintaining its Assumptions?

It might be that the commonly accepted social epistemological approaches to the generation, dissemination, and justification of knowledge (e.g., Hardwig 1985; Kitcher 1990; Longino 2002) are simply insufficient for the analysis of acquiring knowledge from technologies that speak. This is because mainstream approaches regard epistemic processes as socio-cognitive. They ultimately neglect the possibilities of technologies (that speak) participating in knowledge acquisition and belief formation. The question we are left with is how to proceed.

If we wish to evaluate the epistemic roles of technologies that speak in existing social-epistemic terms, we face two options. The first option is rejecting social epistemology's assumptions about the difference between human and non-human agency. Unfortunately, this option will likely lead to a complete revision of fundamental concepts, such as trust, testimony, knowledge, and other concepts currently treated by social epistemologists as anthropocentric.[10]

The second option is introducing new concepts and methods for evaluating the epistemic roles of technologies in current social-epistemic terms. In this option, the new concepts and methods are consistent with fundamental concepts commonly used within mainstream social epistemology. It expands mainstream social epistemology rather than a fundamental revision of it.

While scholars who thought of the examples in §3.3 favor the first option, I favor the second option. Maintaining a distinction between humans and technologies over issues such as agency and morality and keeping concepts such as knowledge and testimony as human-centred is what I believe social epistemology is about: the humane perspective of social knowledge. It is just that our society now has technologies that speak, too. Therefore, in the next section, I offer a concept for analyzing knowledge acquired from technologies in a way that does not constitute a contradiction with received assumptions.

4.2. Knowledge from Natural Language Technologies

Gelfert, in his book *Introduction to Testimony,* offers a category of computer-generated belief:

> it is now entirely conceivable that humans can carry on 'conversations' with computers … which mimick the experience we would have if we were to email back and forth with a (perhaps not overly enthusiastic) human operator. (Gelfert 2014: 27–8)

What about a conversation with a technological artifact enacted via natural language rather than indicators? Taking a cue from Gelfert, I suggest deepening the distinction between testimony-based beliefs and instrument-based beliefs by adding a new category. This new category can encompass the agency-status of the originator of the belief *and* the kind of content that itself eventually becomes a belief. I suggest this third category is *technology-based beliefs*.

---

[10]For discussions about social epistemology as anthropocentric, see Humphreys (2009) and Freiman (2014, 2021).

Table 1. Technology-Based Beliefs: Sources of Knowledge and Content Type.

Epistemic category for the source of the beliefs	Agency of the source of knowledge	Content type
Testimony-based beliefs	Human	Natural language
Instrument-based beliefs	Non-human	Indicators
Technology-based beliefs	Non-human	Natural language

The concept of instrument-based belief correctly captures that the source of the belief, whether perceptual or inferential, is non-human. Similarly, the concept of technology-based belief rests upon the assumption that the source of the content is non-human too. Both categories share that the source of the belief is non-human.

Additionally, the concept of testimony-based belief correctly captures that the testimony is delivered in natural language propositions. Likewise, the concept of technology-based beliefs rests upon the assumption that the non-human delivers propositions in natural language. The 'technology-based belief' solution does not assume that the technology that speaks has a human-like agency. Both categories share that the output content is delivered in natural language (Table 1).

5. Conclusion

Recall Alvin Goldman's question from the beginning of the millennium that was re-raised at the opening of this paper. How will the field of epistemology change in light of technological developments in the means of communication? As conversational AIs become more common, social and ethical challenges become common, too. These days, it is intuitive to say that some beliefs that we acquire from communication technologies qualify as knowledge. It is also fair to assume that this trend is expected to grow.

Despite the dire need to analyze knowledge acquired from these technologies, there is no concept to enable such an analysis. The traditional field of epistemology offers several accounts for knowledge from instruments. However, as argued, these accounts are not suitable for analyzing knowledge from technologies that speak. At the same time, the field of social epistemology offers the concept of testimony to account for acquiring knowledge in the form of natural language. Nevertheless, this concept is not suitable, too, since it assumes that the testifier is human. How can we, for example, epistemically analyze problems such as amplifying sexism and racism, and spreading misinformation by these devices?

To fill the lacuna of analysis of knowledge acquisition from technologies that interact with natural language, this paper suggested the concept of technology-based beliefs. It enjoys the best of all worlds: the agency of the source of knowledge is non-human, and the content is delivered in natural language. The proposed concept can encompass what is currently missing from the field of social epistemology – acquiring knowledge from technologies that speak in natural language.[11]

References

Adewumi T., Li'wicki F. and Liwicki M. (2022). 'State-of-the-art in Open-domain Conversational AI: A Survey.' arXiv preprint. arXiv:2205.00965.

Adler J. (2014) [2006]. 'Epistemological Problems of Testimony.' In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2014/entries/testimony-episprob.

Alvarado R. (2022a). 'What Kind of Trust Does AI Deserve, if Any?' *AI and Ethics*. https://link.springer.com/content/pdf/10.1007/s43681-022-00224-x.pdf.

Alvarado R. (2022b). 'Computer Simulations as Scientific Instruments.' *Foundations of Science* **27**(3), 1183–205.

Baird D. (2004). *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley, CA: University of California Press.

Blake A. (2019). 'Amazon's Alexa Suggests 'Kill Yourself' While Reading From Vandalized Wikipedia Entry.' *The Washington Times*, 26 December. https://www.washingtontimes.com/news/2019/dec/26/amazons-alexa-suggests-kill-yourself-while-reading.

Bloor D. (1999). 'Anti-Latour.' *Studies in the History and Philosophy of Science* **30**(1), 81–112. doi: 10.1016/S0039-3681(98)00038-7.

Boland, H. (2020). 'Amazon's Alexa Accused of Spreading 'Anti-Semitic Conspiracy Theories'.' *The Telegraph*, 26 November. https://www.telegraph.co.uk/technology/2020/11/26/amazons-alexa-fire-spreading-antisemitic-websites-conspiracy.

Brown T. *et al.* (2020). 'Language Models are Few-shot Learners.' *Advances in Neural Information Processing Systems* **33**, 1877–901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Bryson J.J. (2018). *AI & Global Governance: No One Should Trust AI*. United Nations University, Centre for Policy Research, https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html.

Carter J.A. and Pritchard D. (2010). *The Epistemology of Testimony*. Oxford Bibliographies Online: Philosophy.

Coady C.A.J. (1992). *Testimony: A Philosophical Study*. Oxford: Oxford University Press.

Coeckelbergh M. (2012). 'Can we Trust Robots?' *Ethics and Information Technology* **14**(1), 53–60. doi: 10.1007/s10676-011-9279-1.

Collins H.M. (2010). 'Humans not Instruments.' *Spontaneous Generations: A Journal for the History and Philosophy of Science* **4**(1), 138–47. https://spontaneousgenerations.library.utoronto.ca/index.php/SpontaneousGenerations/article/download/11354/11220.

Collins H.M. and Kusch M. (1998). *The Shape of Actions: What Humans and Machines Can Do*. Cambridge, MA: MIT Press.

Collins H.M. and Pinch T. (1993). *The Golem: What Everyone Should Know About Science*. Cambridge: Cambridge University Press.

Comesaña J. (2010). 'Evidentialist Reliabilism.' *Noûs* **44**(4), 571–600. doi: 10.1111/j.1468-0068.2010.00748.x.

Dahl E.S. (2018). 'Appraising Black-boxed Technology: The Positive Prospects.' *Philosophy & Technology* **31**(4), 571–91. doi: 10.1007/s13347-017-0275-1.

De Boer B., Te Molder H. and Verbeek P.P. (2018). 'The Perspective of the Instruments: Mediating Collectivity.' *Foundations of Science* **23**(4), 739–55.

Devlin J., Chang M.W., Lee K. and Toutanova K. (2018). 'Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding.' arXiv preprint. arXiv: 1810.04805.

Faulkner P.R. (2018). 'Collective Testimony and Collective Knowledge.' *Ergo* **5**(4), 103–26.

Floridi L. and Chiriatti M. (2020). 'GPT-3: Its Nature, Scope, Limits, and Consequences.' *Minds & Machines* **30**, 681–94. doi: 10.1007/s11023-020-09548-1.

Freiman, O. (2014). 'Towards the Epistemology of the Internet of Things: Techno-Epistemology and Ethical Considerations Through the Prism of Trust.' *International Review of Information Ethics* **22**(2), 6–22. doi: 10.29173/irie124.

Freiman, O. (2021). *The Role of Knowledge in the Formation of Trust in Technologies*. PhD dissertation, Bar-Ilan University.

Freiman, O. (2022). 'Making Sense of the Conceptual Nonsense 'Trustworthy AI'.' *AI and Ethics*. doi: 10.1007/s43681-022-00241-w.

Freiman O. and Geslevich Packin N. (2022). 'Artificial Intelligence Products Cannot be Moral Agents.' *Toronto Star*, 7 August. https://www.thestar.com/opinion/contributors/2022/08/07/artificial-intelligence-products-cannot-be-moral-agents-the-tech-industry-must-be-held-responsible-for-what-it-develops.html.

Freiman O. and Miller B. (2020). 'Can Artificial Entities Assert?' In S. Goldberg (ed.), *The Oxford Handbook of Assertion*, pp. 415–36. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780190675233.013.36.

Fricker E. (1995). 'Critical Notice: Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony.' *Mind* **104**(414), 393–411.

Fricker E. (2002). 'Trusting Others in the Sciences: A Priori or Empirical Warrant?' *Studies in History and Philosophy of Science Part A* **33**(2), 373–83. doi: 10.1016/S0039-3681(02)00006-7.

Fricker E. (2015). 'How to Make Invidious Distinctions Amongst Reliable Testifiers.' *Episteme* **12**(2), 173–202. doi: 10.1017/epi.2015.6.

Fricker E. and Cooper D.E. (1987). 'The Epistemology of Testimony.' *Proceedings of the Aristotelian Society, Supplementary Volumes* **61**, 57–106.

Fu T., Gao S., Zhao X., Wen J.R. and Yan R. (2022). 'Learning Towards Conversational Ai: A Survey.' *AI Open* **3**, 14–28.

Gelfert A. (2014). *A Critical Introduction to Testimony*. London: A&C Black.

Gelfert A. (2018). *Testimony*. London: Routledge.

Giere R.N. (2006). *Scientific Perspectivism*. Chicago, IL: University of Chicago Press.

Goldberg S.C. (2012). 'Epistemic Extendedness, Testimony, and the Epistemology of Instrument-based Belief.' *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* **15**(2), 181–97. doi: 10.1080/13869795.2012.670719.

Goldman A.I. (2000). 'Telerobotic Knowledge: A Reliabilist Approach.' In K. Goldberg (ed.), *The Robot in the Garden*, pp. 126–42. Cambridge, MA: MIT Press.

Goldman A.I. (2011). 'Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Troubles, Reliabilism's Rescue Package.' In T. Dougherty (ed.), *Evidentialism and its Discontents*. Oxford: Oxford University Press.

Green C.R. (2006). *The Epistemic Parity of Testimony, Memory, and Perception*. PhD dissertation, University of Notre Dame.

Green C.R. (2008). 'Epistemology of Testimony.' In *Internet Encyclopedia of Philosophy*. https://www.iep.utm.edu/ep-testi.

Hacking I. (1985). 'Do We See Through a Microscope?' In P.M. Churchland and A.C. Hooker (eds), *Images of Science: Essays on Realism and Empiricism, (with a reply from Bas C. van Fraassen)*, pp. 132–52. Chicago, IL: University of Chicago Press.

Hardwig J. (1985). 'Epistemic Dependence.' *Journal of Philosophy* **82**(7), 335–49. doi: 10.2307/2026523.

Harwell D., Tiku N. and Oremus W. (2022). 'Stumbling with their Words, Some People let AI do the Talking.' *Washington Post*, 10 December. https://www.washingtonpost.com/technology/2022/12/10/chatgpt-ai-helps-written-communication.

Henderson D., Horgan T. and Potrč M. (2007). 'Transglobal Evidentialism-Reliabilism.' *Acta Analytica* **22**(4), 281–300. doi: 10.1007/s12136-007-0015-8.

Humphreys P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

Humphreys P. (2009). 'Network Epistemology.' *Episteme* **6**(2), 221–9. doi: 10.3366/E1742360009000653.

Hurst L. (2022). 'ChatGPT: Why the Human-like AI Chatbot Suddenly Has Everyone Talking.' *EuroNews*, 14 December. https://www.euronews.com/next/2022/12/14/chatgpt-why-the-human-like-ai-chatbot-suddenly-got-everyone-talking.

Ihde D. (1991). *Instrumental Realism: The Interface Between Philosophy of Science and Philosophy of Technology*. Indianapolis, IN: Indiana University Press.

Kitcher P. (1990). 'The Division of Cognitive Labor.' *Journal of Philosophy* **87**(1), 5–22.

Knorr-Cetina K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kusch M. (2002). *Knowledge by Agreement: The Programme of Communitarian Epistemology*. Oxford: Oxford University Press.

Lackey J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

Lackey J. (ed.) (2014). 'A Deflationary Account of Group Testimony.' In *Essays in Collective Epistemology*, pp. 64–94. Oxford: Oxford University Press.

Latour B. (1986). 'Visualization and Cognition: Thinking with Eyes and Hands.' *Knowledge and Society* **6**, 1–40.

Laudan L. (1981). 'A Confutation of Convergent Realism.' *Philosophy of Science* **48**, 19–48.

Lehrer K. (1990). *Theory of Knowledge*. Boulder, CO: Westview Press.

Lehrer K. (1995). 'Knowledge and the Trustworthiness of Instruments.' *The Monist* **78**(2), 156–70. doi: 10.5840/monist199578216.

Lehrer K. (2000). *Theory of Knowledge*. 2nd edition. Boulder, CO: Westview Press.

Lehrer K. (2003). 'Coherence, Circularity and Consistency: Lehrer Replies.' In E.J. Olsson (ed.), *The Epistemology of Keith Lehrer*, pp. 309–56. Amsterdam: Kluwer.

Lipton P. (1998). 'The Epistemology of Testimony.' *Studies in History and Philosophy of Science Part A* **29** (1), 1–31. doi: 10.1016/S0039-3681(97)00022-8.

Longino H. (2002). *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

Lynch M. (1994). 'Representation is Overrated: Some Critical Remarks About the Use of the Concept of Representation in Science Studies.' *Configurations* **2**(1), 137–49. doi: 10.1353/con.1994.0015.

Matilal B.K. and Chakrabarti A. (eds) (1994). *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*. Dordrecht: Springer.

Meaker M. (2019). 'How Digital Virtual Assistants Like Alexa Amplify Sexism.' *Medium OneZero*, 10 May. https://onezero.medium.com/how-digital-virtual-assistants-like-alexa-amplify-sexism-8672807cc31d.

Millar A. (2009). 'Knowledge and Recognition.' In *The Nature and Value of Knowledge: Three Investigations*, pp. 91–190. Oxford: Oxford University Press.

Miller B. (2015). 'Why (Some) Knowledge is the Property of a Community and Possibly None of Its Members.' *Philosophical Quarterly* **65**(260), 417–41.

Miller B. and Freiman O. (2020). 'Trust and Distributed Epistemic Labor.' In J. Simon (ed.), *The Routledge Handbook on Trust and Philosophy*, pp. 341–53. London: Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315542294-26/trust-distributed-epistemic-labor-boaz-miller-ori-freiman.

Mollman S. (2022). 'ChatGPT Gained 1 Million Users in Under a Week. Here's Why the AI Chatbot is Primed to Disrupt Search as we Know it.' *Yahoo! Finance*, 9 December. https://finance.yahoo.com/news/chatgpt-gained-1-million-followers-224523258.html.

Neges (Kletzl) S. (2014). 'Scrutinizing Thing Knowledge.' *Studies in History and Philosophy of Science Part A*, **47**, 118–123. doi: 10.1016/j.shpsa.2014.06.002.

Neges (Kletzl) S. (2018). 'Instrumentation. A Study in Social Epistemology.' PhD dissertation, University of Vienna.

Nickel P.J. (2013). 'Artificial Speech and its Authors.' *Minds and Machines* **23**(4), 489–502.

Olesen F. (2012). 'Scientific Objectivity and Postphenomenological Perception.' *Foundations of Science* **17** (4), 357–62. doi: 10.1007/s10699-011-9241-z.

Olsson E. (2017). 'Coherentist Theories of Epistemic Justification.' In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2017/entries/justep-coherence.

OpenAI (2022). 'ChatGPT: Optimizing Language Models for Dialogue.' 30 November. https://openai.com/blog/chatgpt/.

Pitt J.C. (2007). 'Speak to Me.' *Metascience*, **16**, 51–59. doi: 10.1007/s11016-006-9070-9.

Pitt J.C. (2010). 'It's not About Technology.' *Knowledge, Technology and Policy* **23**(3–4), 445–54. doi: 10.1007/s12130-010-9125-5.

Pritchard D. (2004). 'The Epistemology of Testimony.' *Philosophical Issues* **14**(1), 326–48. doi: 10.1111/j.1533-6077.2004.00033.x.

Rieder G., Simon J. and Wong P.H. (2020). 'Mapping the Stony Road Toward Trustworthy AI: Expectations, Problems, Conundrums.' In *Machines We Trust: Perspectives on Dependable AI*. Cambridge, MA: MIT Press.

Russell S.J. & Norvig P. (2021). *Artificial Intelligence: A Modern Approach*. Fourth Edition. London: Pearson.

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

Shapin S. (1994). *A Social History of Truth*. Chicago, IL: University of Chicago Press.

Smart P.R. (2017). 'Extended Cognition and the Internet: A Review of Current Issues and Controversies.' *Philosophy and Technology* **30**(3), 357–90.

Sosa E. (2006). 'Knowledge: Instrumental and Testimonial.' In J. Lackey and E. Sosa (eds), *The Epistemology of Testimony*, pp. 116–23. Oxford: Oxford University Press.

Strubell E., Ganesh A. and McCallum A. (2019). 'Energy and Policy Considerations for Deep Learning in NLP.' arXiv preprint. arXiv: 1906.02243.

Symons J. and Alvarado R. (2019). 'Epistemic Entitlements and the Practice of Computer Simulation.' *Minds and Machines* **29**(1), 37–60. https://link.springer.com/article/10.1007/s11023-018-9487-0#Sec8.

Thoppilan R. *et al.* (2022). 'Lamda: Language Models for Dialog Applications.' arXiv preprint. arXiv: 2201.08239.

van Fraassen B.C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

Verbeek P.P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, PA: Penn State Press.

Waelbers K. and Briggle A. (2010). 'Three Schools of Thought on Freedom in Liberal, Technological Societies.' *Techné: Research in Philosophy and Technology* **14**(3), 176–93. doi: 10.5840/techne201014320.

Ori Freiman is a Post-Doctoral Fellow at McMaster University's Digital Society Lab. He is researching the responsible implementation of emerging technologies, focusing on the democratic implications of Central Bank Digital Currencies, the shaping of national AI policies, and the development of 'Responsible AI'.